THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

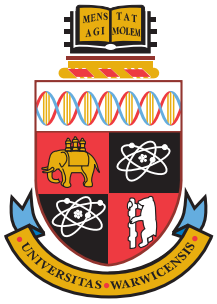
A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/3127

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.



Characterisation of bacteriophages that infect *Acaryochloris*

by

Yi-Wah Chan

Thesis
Submitted to the University of Warwick
for the degree of
Doctor of Philosophy

Supervisors: Dr Martha R. J. Clokie, Dr Anna L. Whitworth and Prof. David J. Scanlan

MOAC Doctoral Training Centre
January, 2010

THE UNIVERSITY OF WARWICK

Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to thank my wonderful supervisor, Martha. She has supported and encouraged me throughout my PhD and one could not ask more of her. I wish her and her family all the best for the future.
I thank Dave for supervising, his keen eye and all his good advice. I would also like to thank Anna for her supervision and wish her good health for the future. I would like to thank Andy, for his sense of humour, his support in teaching me everything I needed to know, his perserverance and his utmost patience. I would like to thank my supervisory committee: Nick, Pat and Colin for their advice and support and all my friends and colleagues at Biology, Chemistry and MOAC for making the last four years a fun experience.
I would like to thank my collaborators Sam, Anja, Remus, Peter, Wolfgang, Anthony and Michelle. I would like to thank Svetla and Jane for microscopy training and Martin for help with LabView. A big thank you to Cerith for the endless volumes of ASW, Ken for all his hard work, Marcus and Lee for making all the parts for my Faraday cage and all the people at Biology, Chemistry and MOAC who made everything run smoothly. I would also like to thank the EPSRC for funding, NERC for funding the genome sequencing and Berthold for performing the sequencing.
I would like to thank Hugh for being awesome, Fiona for being brilliant, Jana for being Jana, the gang for an unforgettable time, Ioana for hei from

the beginning, Antony for his friendship (and all the programs he wrote for me!!! seriously.), Laha for listening and Mike for random emails... I would also like to thank Ben and Jaytry without whom, I would not have chosen this project.

Finally, I would like to thank my family for their support: my parents, my brother, Yi, my sister, Yi, Monty, Seb and Millie.

^^ y.

Declaration

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work in this thesis has been undertaken by myself except where otherwise stated. None of the material in this thesis has been previously published. All quoted text remains the copyright of the original attributed author. I have taken the liberty of correcting typographical errors, spelling and grammar where necessary. All trademarks are acknowledged.

Every care was taken when referencing websites to ensure only major websites concerned with a particular subject were referenced to minimise the possibility of them subsequently becoming invalid. All web addresses were valid as of August, 2009. Work included in Chapter 2 was published in (Chan et al., 2007) and the publication is included in Appendix B.

Yi-Wah Chan

January, 2010

Abbreviations

A	adenine
ABC	ammonium bicarbonate
ACN	acetonitrile
Anc_*rnt*	ancestral RNase T
APC	allophycocyanin
APS	ammonium persulfate
ASW	artifical seawater
BCA	bicinchoninic acid
BLAST	basic local alignment search tool
C	cytosine
CAMERA	Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis
carbpep	carboxypeptidase
CCMEE	Culture Collection of Microorganisms from Extreme Environments
cDNA	complementary DNA
Chl	chlorophyll
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSR	cell surface receptor
CV	cyclic voltammogram
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNK	deoxynucleoside kinase
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
*Ec*Exo	*Escherichia coli* exonuclease
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
End	endonuclease

FcTMA	ferrocenylmethyltrimethylammonium hexafluorophosphate
FEN	flap endonuclease
FSC	forward scatter
G	guanine
GmbH	Gesellschaft mit beschränkter Haftung
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IgG	immunoglobulin G
IT	current-time
LC	liquid chromatography
LGT	lateral gene transfer
LSCM	laser scanning confocal microscopy
Ltd.	limited
maj	major
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight
MBIC	Marine Biotechnology Institute Culture Collection
min	minor
MOI	multiplicity of infection
mRNA	messenger RNA
MS	mass spectrometry
mtase	methyltransferase
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
NIR	near infrared radiation
nr	non-redundant
NTPase	nucleoside triphosphatase
OD	optical density
ORF	open reading frame
ϕ*rnt*	phage RNase T
PAGE	polyacrylamide gel electrophoresis
PAM	pulse amplitude-modulated
PAR	photosynthetically active radiation
PBP	phycobiliprotein
PBS	phycobilisome
PBS-T	phosphate buffered saline with 0.1 % Tween
PC	phycocyanin
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFGE	pulsed-field gel electrophoresis
p.f.u.	plaque forming unit
PHIRE	PHage *In silico* Regulatory Elements
PIM	percent identity matrix
pol	polymerase
pri	primase
PS	photosystem

qPCR	quantitative (real time) polymerase chain reaction
RBS	ribosomal binding site
RE	reference electrode
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
*Sa*FEN	*Staphylococcus aureus* 5'-3' exonuclease
SECM	scanning electrochemical microscopy/microscope
SEM	scanning electron microscopy/microscope
SDS	sodium dodecyl sulphate
SSC	side scatter
T	thymine
TBE	Tris/Borate/EDTA
TE	Tris/EDTA
TEM	transmission electron microscopy/microscope
TEMED	tetramethylethylenediamine
TerL	terminase large subunit
TMP	tape measure protein
Tris	trishydroxymethylaminomethane
tRNA	transfer RNA
UA	uranyl acetate
UME	ultramicroelectrode
UPW	ultrapure water
UV	ultraviolet
VSP	virion structural protein
WE	working electrode

Summary

The cyanobacterium *Acaryochloris marina* was isolated in 1996 and solved a 50 year old mystery as to the origin of the pigment chlorophyll d, which was thought to be a pigment of red algae or a breakdown product of the universal chlorophyll, chlorophyll a. Over the next decade, new *Acaryochloris* spp. were isolated from all over the world as the genus received international interest from the scientific community, with the majority of research directed towards understanding the mechanisms of photosynthesis of this uniquely pigmented cyanobacterium, using *A. marina* as the model organism of the genus. During this project, characterisation of different aspects of photosynthesis in *Acaryochloris* spp. was performed including an investigation of pigment adaptation and composition and the growth and characterisation of *A. marina* biofilms. However, the main focus of the thesis concerns the isolation and characterisation of cyanophages A-HIS1 and A-HIS2, which infect *A. marina* as a basis to investigate and understand the impact of phage on host physiology in this new model system. A-HIS1 and A-HIS2 were characterised by their morphology, growth behaviour and genomes. Experiments were designed and implemented to investigate interactions between the phages and host. Interestingly, an analysis of novel genes in these phages revealed a surprising evolutionary history of phages A-HIS1 and A-HIS2 providing new insights into the origin of DNA polymerase γ, which is found only in the mitochondria of eukaryotes.

Chapter 1

Introduction

1.1 Cyanobacteria

Cyanobacteria are unicellular photosynthetic prokaryotes responsible for the aerobic atmosphere we breathe today (Kirschvink et al., 2000; Kump, 2008). *Synechococcus* spp. and *Prochlorococcus* spp. are the dominant marine cyanobacteria, which have been shown to be significant contributors to global primary productivity (Waterbury et al., 1986; Liu et al., 1997; Partensky et al., 1999) and oceanic carbon stores (Goericke and Welschmeyer, 1993; Li, 1995; Veldhuis et al., 1997). This thesis is primarily concerned with *Acaryochloris*, a genus of cyanobacteria, which is characterised by its unique pigmentation, life-style and habitats. This chapter introduces the *Acaryochloris* genus, bacteriophages, the hypotheses which motivated this thesis and the aims of this work.

1.2 *Acaryochloris* spp.

1.2.1 History and habitats

In 1996, Miyashita et al. reported the discovery of an oxygenic photosynthetic prochlorophyte-like prokaryote predominantly containing the green pigment chlorophyll (Chl) *d*. The cyanobacterium was discovered living in

a suspension of algae squeezed from a colonial ascidian (or sea squirt) *Lissoclinum patella* off the coast of the Palau Islands in the western Pacific Ocean (Miyashita et al., 1996). The prokaryote was named *Acaryochloris marina* Miyashita et Chihara gen. et sp. nov. strain MBIC11017, which is maintained at the Marine Biotechnology Institute Culture collection.

It has since been demonstrated that *Acaryochloris* spp. are widespread; they have been found inside or beneath a number of Didemnidae genera (Miyashita et al., 1996; Kühl et al., 2005; Larkum and Kühl, 2005), on algal species (Murakami et al., 2004; Ohkubo et al., 2006), free-living (Miller et al., 2005), as granite endolithobionts (de los Ríos et al., 2007) and most recently associated with columnar stromatolites (Goh et al., 2009).

Interestingly, the *Acaryochloris* genus of cyanobacteria is thought to have been encountered 53 years prior to its discovery in 1996 by the chemists Winston M. Manning and Harold H. Strain in 1943, who presented the discovery of Chl *d* as a previously undiscovered green pigment of red algae (Manning and Strain, 1943). It was later suggested by Holt and Morley that Chl *d* could be a by-product induced during pigment extraction (Holt and Morley, 1959; Holt, 1961) and the subject of Chl *d* and its origins were left undisturbed until the discovery of the prokaryote.

In 2004, Murakami et al. identified and isolated *Acaryochloris* sp. strain Awaji living epiphytically on the thalli of the red alga *Ahnfeltiopsis flabelliformis* (Murakami et al., 2004). They also observed that *Acarychloris* spp. were found on other marine Rhodophyta, namely *Callophyllis japonica* and *Carpopeltis prolifera*. Ohkubo et al. have since shown that *Acaryochloris* spp. also live on green and brown algae (Ohkubo et al., 2006). Finally, the puzzle of Chl *d* and its origins were resolved (Larkum and Kühl, 2005).

1.2.2 Biofilms: symbiont or epibiont?

Nearly a decade after the isolation of *A. marina* from within a colonial didemnid ascidian as presented by Miyashita et al. (Miyashita et al., 1996), Larkum and Kühl discussed the true habitat and nature of *Acaryochloris* spp. (Larkum and Kühl, 2005), specifically, whether they lived as symbionts within or as epibionts beneath ascidians. Earlier in 2005, Kühl et al. discussed the niche-nature of *Acaryochloris* spp. characterised by their habitats, in that they exploit near infrared radiation (NIR) by containing mainly Chl *d* which absorbs light in the far red-light region above 680 nm (Kühl et al., 2005).

Kühl et al. presented ample evidence in their microphotometric survey of didemnid ascidians (collected from the North reef flat of Heron Island, Great Barrier Reef) indicating *Acaryochloris* spp. form biofilms on the underside of didemnid ascidians as opposed to being symbiotic within them. Kühl et al. also used fibre-optic microprobe measurements on the ascidian *Diplosoma virens* to show that visible light was attenuated by the ascidian tissue, providing strong data to support their hypothesis that *Acaryochloris* spp. live beneath ascidians.

To add to the debate on the true habitat of *Acaryochloris* spp. *Acaryochloris* sp. strain CCMEE5410[1] was discovered free-living on an epilithic microbial mat on the southern shore of the Salton Sea, a eutrophic and highly polluted, hypersaline lake (Miller et al., 2005). Larkum and Kühl argued that the presence of *Acaryochloris* spp. on the underside of red algae (Murakami et al., 2004), which strongly deplete visible light and enhance NIR and the supporting evidence of the microphotometric survey not only resolved the issue of the source of Chl *d* in the study by Manning and Strain (Manning and Strain, 1943), but also showed that these *Acaryochloris* spp. probably

[1]Maintained in the University of Oregon Culture Collection of Microorganisms from Extreme Environments.

originally lived in these types of light-depleted habitat. However, Larkum and Kühl could not ignore that the nature of the ascidian habitat implied that there was a possibility that *Acaryochloris* spp. could be ingested by the ascidians. To date, most evidence suggests that *Acaryochloris* spp. are free-living cyanobacteria which may live epibiontically on metazoans, where light is enriched in NIR.

1.2.3 Chlorophyll *d*

The main pigment of the *Acaryochloris* genus is Chl *d*; they only possess a small amount of Chl *a* (Fig. 1.1(A)), the universal chlorophyll, which appears as the dominant chlorophyll in all other oxygenic phototrophs (Scheer, 1991). Chl *a* only makes up 3 % of the chlorophyll in *A. marina* (Miyashita et al., 1996). There is very little structural difference between Chl *a* and Chl *d*; the C3 group (Elmsford et al., 1979) for Chl *d* is a formyl group, but in the case of Chl *a*, the C3 group is a vinyl group (Fig. 1.1(A)).

Absorption spectra in the visible/infra-red range of light from 350–800 nm are commonly used to compare chlorophylls, which are generally characterised by peak wavelengths or wavelength bands. It turns out that the seemingly small difference in the chlorophyll C3-group which separates Chl *a* from Chl *d* has a dramatic effect on the spectral peak referred to as the Q_y-absorption peak[2] (Shipman et al., 1976), which occurs in the range 660–750 nm (Schiller et al., 1997; Larkum and Kühl, 2005), called the Q_y band. This is clearly seen on comparison of whole cell absorption spectra of the cyanobacteria *Synechocystis* sp. PCC6803 and *A. marina* (Fig. 1.1(B)), which possess mainly Chl *a* and Chl *d*, respectively. Chl *a* has its Q_y-peak at 680 nm, compared to Chl *d*, where the Q_y-peak is red-shifted to 716 nm. Alternatively, spectral differences can be observed by examining the char-

[2] Q_y refers to a transition from the ground electronic state to the first excited singlet electronic state of for example Chl *a*, which is responsible for the visible absorption spectrum of Chl *a* in the red and orange regions.

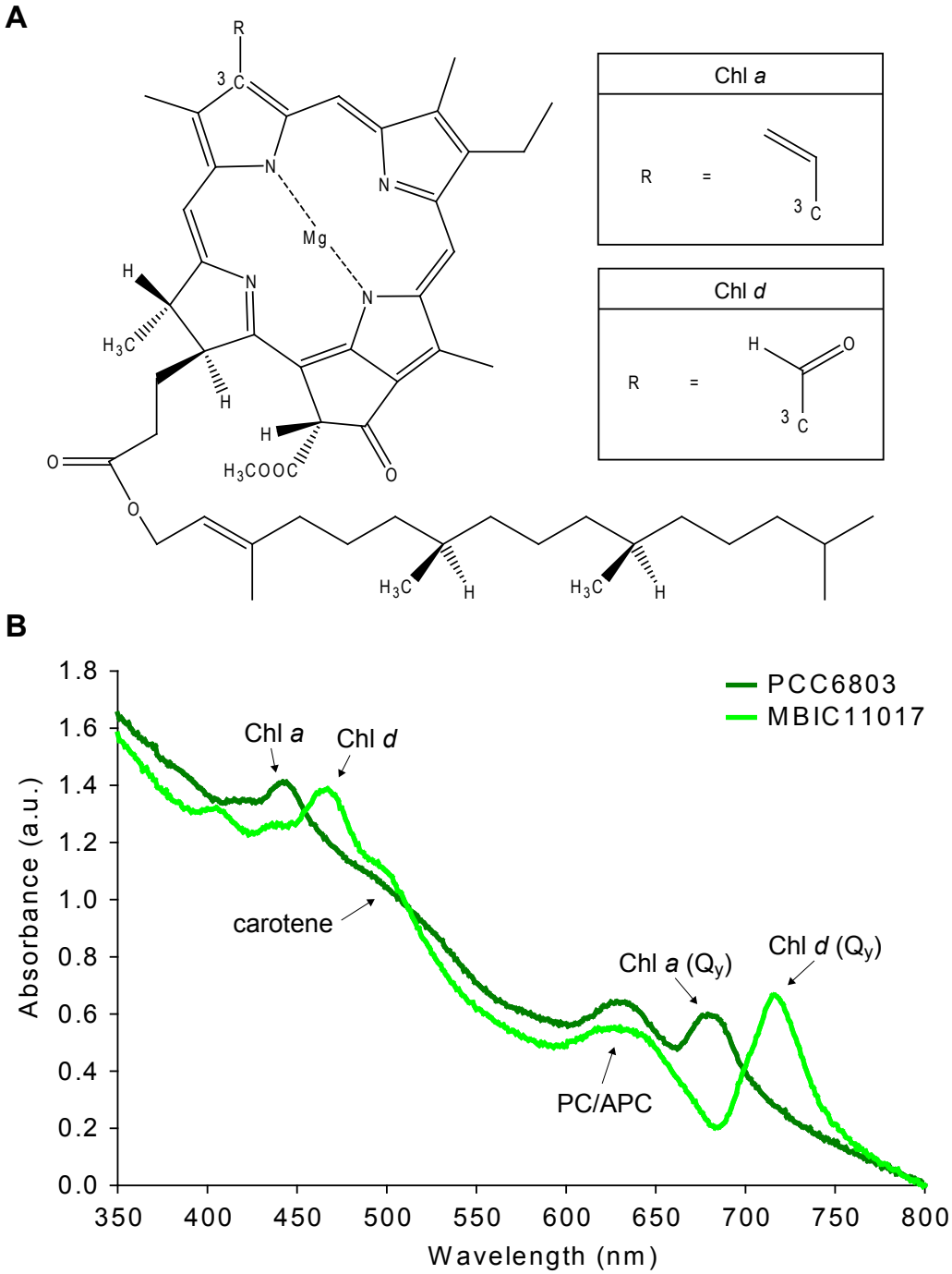


Fig. 1.1: Chlorophylls *a* and *d*. (A) Molecular structure of Chl *a* and Chl *d* - the structure of each is determined by the respective C3 functional group, R. (B) Whole cell absorbance spectra of *Synechocystis* sp. PCC6803 and *A. marina* MBIC11017. Phycobiliproteins: phycocyanin = PC, allophycocyanin = APC.

acteristics of the chlorophylls when extracted from cells by solvents such as methanol (Larkum and Kühl, 2005), where essentially the main differences are that in solvents, the spectra are blue-shifted and light-scattering caused by the cells is removed.

The unique properties of Chl *d* of the *Acaryochloris* genus have consequently led to the worldwide interest in this genus. Importantly, Chl *d* extends photosynthetically active radiation or PAR (Gates, 1980). PAR refers to the spectral range 400–700 nm, which was coined to describe the light useful to photosynthetic organisms for photosynthesis.

1.2.4 Phycobiliproteins and photosystems

Phycobilisomes (PBS) are pigment-protein complexes which absorb light and transfer the light energy to chlorophyll in the photosystems. PBS are composed of smaller subunits called phycobiliproteins, which are part bilin (pigment) and part protein (Glazer, 1982). Most cyanobacteria use a combination of PBS and the photosystems I and II (PSI and PSII respectively) to harvest light and some are able to chromatically adapt and alter their pigment ratios depending on the light conditions (Tandeau de Marsac, 1977; Palenik, 2001).

A. marina does not appear to contain PBS, but instead possesses biliprotein aggregates, which form rod-shaped structures, which were thought to be some form of precursor of PBS (Marquardt et al., 1997). The biliprotein aggregates were shown to be composed of two types of biliproteins: phycocyanin (PC) and allophycocyanin (APC) and the biliproteins are thought to be attached to the PS II complexes of *A. marina* (Hu et al., 1999). The contribution of these two pigments to the whole cell absorption spectrum of *A. marina* can be observed in Fig. 1.1(B) from 590–690 nm. In comparison to other aspects of *A. marina*, the photosystems have by far been the subject of most research. This has focused on the chlorophylls which are

associated with PSI and PSII, namely the localization of Chl *a* or Chl *d* in the photosynthesis apparatus (Hu et al., 1998; Mimuro et al., 1999, 2004; Chen et al., 2005; Renger and Schlodder, 2008).

1.3 Bacteriophages

Bacteriophages (phages) are viruses which infect bacteria. They are the most abundant biological entities on Earth and have a direct influence on bacterial populations, since they play key environmental roles in biogeochemical cycling, population genetics and bacterial evolution through lateral gene transfer (LGT). The sheer abundance of bacteriophages and their inextricable relationship with bacteria has put them at the forefront of marine research (Torrella and Morita, 1979; Bergh et al., 1989; Suttle, 2005, 2007).

Of all the phages which have been subjected to scientific study, 96 % of the phages are from the order *Caudovirales* of double-stranded DNA (dsDNA) tailed phages (Loessner and Hagens, 2007), which is comprised of the families *Myoviridae*, *Siphoviridae* and *Podoviridae*. All three families are characterised by icosahedral protein capsids referred to as the phage head, which contains the dsDNA phage genome.

Myoviruses are characterised by a long, contractile tail, whereas siphoviruses have long, non-contractile, flexible tails and podoviruses have short non-contractile tails (Suttle, 2005). To date, all phages which have been isolated on cyanobacteria (cyanophages), belong to the *Caudovirales*. Illustrations of the typical morphologies of the three families of the *Caudovirales* are presented in Fig. 1.2(A/B/C).

Phages are generally regarded as abiotic, non-living and classed as either lytic or temperate depending on their mode of replication. Lytic phages as their name suggests, introduce their genetic material into the host, replicate new phage progeny and then lyse the host cell, releasing the phage

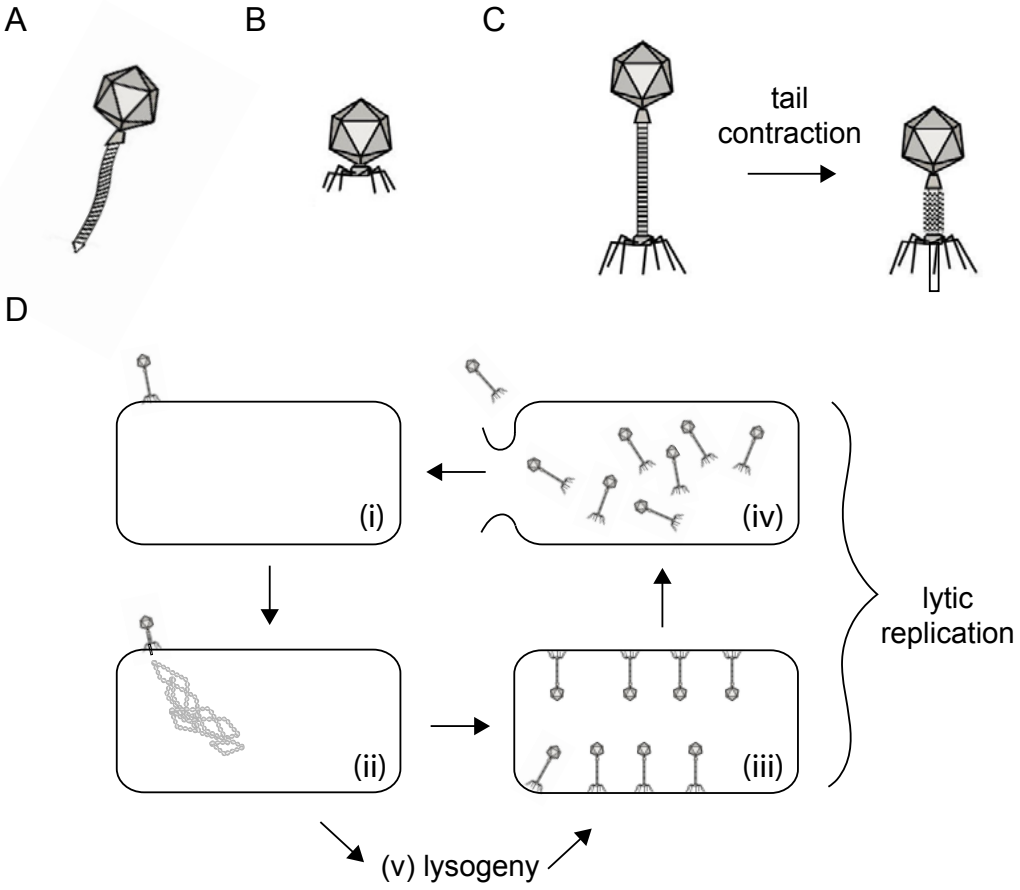


Fig. 1.2: The *Caudovirales* and phage replication. Generalised representation of (A) *Siphoviridae*, (B) *Podoviridae* and (C) *Myoviridae* in native form and with tail contracted exposing tail-pipe and (D) schematic of modes of phage replication; lytic replication involves (i) phage adsorption, (ii) the injection of phage DNA, (iii) the creation of new progeny followed by (iv) lysis at which point the progeny can go on to infect other cells. Alternatively, after (ii), phages may undergo (v) lysogeny by incorporating the phage DNA into the host genome forming a lysogen (bacterium with prophage). The lysogen will either continue to divide or the prophage may be induced by environmental stress and the lysogen may enter the lytic replication cycle (UV light or chemicals such as mitomycin C are typically used in the laboratory to induce prophages) (Abedon, 2008).

progeny. In the lysogenic cycle, after the introduction of phage genetic material to the host cell, the phage genome is integrated as a prophage into the host genome, which will continue to divide until such a time when the prophage is activated either naturally or artificially in the laboratory. There are many other modes of infection which are summarised by Abedon (Abedon, 2008); phage infections may be categorised as chronic, pseudolysogenic, phage restrictive and abortive depending on the viability of phage and bacteria as well as lytic and lysogenic infections, which were previously mentioned. These other forms of infection are not considered here as the phages isolated appear to be lytic.

1.3.1 Cyanophages

Phages have never been isolated on *Acaryochloris* spp. or on any cyanobacterium with a sedentary lifestyle. This includes cyanobacteria which live either symbiotically with or epiphytically on a metazoan (Kühl et al., 2005; Larkum and Kühl, 2005). In recent years, phage biologists (and biologists interested in cyanophage) have turned to genome data in order to understand the relationships between phages and the role they play in the phage-host systems. To date, only eight lytic marine cyanophages have had their genomes sequenced, and all infect either *Synechococcus* spp. or *Prochlorococcus* spp. (or both in the case of the cyanophage Syn9 (Weigele et al., 2007)). These phages are from the *Myoviridae* and *Podoviridae* and are listed in Table 1.1.

Three lytic freshwater cyanophage have also had their genomes sequenced: the podoviruses Pf-WMP3 (Liu et al., 2008) and Pf-WMP4 (Liu et al., 2007) which infect *Phormidium foveolarum* and the myovirus Ma-LMM01 which infects *Microcystis aeruginosa* (Yoshida et al., 2008). So far, no cyanophages of the family *Siphoviridae* have had their genome sequenced.

Table 1.1: Marine cyanophages with sequenced genomes.

Family	Phage	Reference
Myoviridae	S-PM2	(Mann et al., 2005)
	Syn9	(Weigele et al., 2007)
	S-RSM4	(Millard et al., 2009)
	P-SSM2	(Sullivan et al., 2003)
	P-SSM4	
Podoviridae	P-SSP7	
	P60	(Chen and Lu, 2002)
	Syn5	(Pope et al., 2007)

1.3.2 Lateral gene transfer in cyanophages

Lateral gene transfer or LGT (also called horizontal gene transfer) describes processes by which genetic material is transferred between two organisms via any non-vertical transfer from an ancestor. Cyanobacterial phages have been shown to contain photosynthesis-related genes obtained from the host by LGT, for example the *Synechococcus* myovirus S-PM2 encodes the D1 and D2 proteins (amongst other proteins related to photosynthesis) which are core components of the PSII reaction centre (Mann et al. 2005). Similarly, the *Prochlorococcus* cyanophages P-SSM2, P-SSM4 (myoviruses) and P-SSP7 (podovirus) were all found to encode different combinations of photosynthesis-related genes, which include the D1 and D2 proteins and a high-light inducible protein amongst others (Lindell et al., 2004).

The genome data available has led to the growing influx of evidence that phages are key drivers for LGT not only between host organisms but also phages (Mirold et al., 2001; Canchaya et al., 2003). The emerging picture surrounding the evolution of organisms and the viruses which infect them is one of a network of genetic data transfer coupled with a natural selection process, which by its nature is inherently complex and diverse. This poses a significant problem to evolutionary biologists and taxonomists who are concerned with bacteria, as it becomes increasingly clear that the content of any one genome is a historical legacy of genetic interaction with other

organisms which may be marked by opportunistic phage-mediated transfer events (Hendrix et al., 1999).

1.3.3 Phage orchestration of photosynthesis

The first phage shown to contain photosynthesis-related genes was the *Synechococcus* sp. WH7803 phage S-PM2 (Mann et al. 2003). This host-phage system has been well-characterised showing the orchestration of photosynthesis by the phage upon infection of the host. Shortly after, this phenomenon was shown to extend to *Prochlorococcus* phages (Lindell et al. 2004). Quantitative real-time polymerase chain reaction (qPCR) has since been used to elucidate gene expression levels of the phage encoded photosynthesis-related genes in both *Prochlorococcus* and *Synechococcus* phage-host systems (Lindell et al., 2005; Clokie et al., 2006a; Shan et al., 2008). More recently microarray studies have been used to study expression dynamics (Lindell et al., 2007).

The action of S-PM2 on its host has been particularly well-characterised and is considered here in more detail. Clokie et al. have shown that the phage homologue of the photosystem protein D1 encoded by *psbA* was expressed shortly after host cells were infected and that the phage homologue essentially complemented the host transcript levels to maintain the photosynthetic activity of the host as it was before infection up until lysis. Phage S-PM2 also encodes other photosynthesis-related genes like *cpeT*, which may encode a regulator of phycobiliprotein biosynthesis since it is typically found with *cpeR*, which has been shown to be an activator required for the expression of the phycoerythrin operon *cpeBA* in a study of the cyanobacterium *Fremyella diplosiphon* (Cobley et al., 2002). In this case, cpeT is encoded in the phycoerythrin linker-polypeptide operon *cpeCDESTR*.

Recently, Shan et al. demonstrated a modification in the pigment content of *Synechococcus* sp. WH7803 after infection with the phage S-PM2 (Shan

et al. 2008), where an increase in Chl *a* was observed as well as a gradual increase in the size of the PBS which corresponded to an increase in phycoerythrin. Importantly, the transcript level of the *cpe*BA operon in WH7803 was also shown to be upregulated after phage infection. Overall, it has been concluded that the presence of the photosynthesis-related genes in S-PM2 may influence the efficiency with which the host may perform in producing new phage progeny, especially since the replication cycle of S-PM2 in WH7803 is lengthy at 9 hours before lysis (Wilson et al. 1996), during which light conditions may change in the natural environment.

1.4 Thesis aims, motivations and outline

The main aims of this study were to isolate and characterise cyanophages which infect *A. marina* and to determine their impact on the photosynthesis of their host cyanobacterium. To determine how phage impact the physiology of *A. marina* it is necessary to understand the physiology of their uninfected hosts and this is the focus of **Chapter 2** and **Chapter 6**. **Chapter 2** details how *Acaryochloris* spp. strains MBIC11017 and CCMEE5410 adapt their pigment composition under different growth conditions. *Acaryochloris* spp. have been shown to live as epibionts as well as free-living in a variety of different habitats which are exposed to different light conditions (Miyashita et al., 1996; Murakami et al., 2004; Miller et al., 2005).

The work in this chapter was motivated by initial observations made in the laboratory which showed that the two strains exhibited different pigmentations under different culturing conditions of stock cultures. Furthermore, in 2002, Chen et al. noted that under 100 μmol photons $m^{-2}s^{-1}$ the PBP content of strain MBIC11017 was decreased compared to growing at 10–30 μmol photons $m^{-2}s^{-1}$ (Chen et al., 2002), but no further investigation was performed to assess the effect of light irradiance on the pigmentation of strain MBIC11017 in terms of an ability to adapt to different light quan-

tities. Concurrently, strain CCMEE5410 was newly isolated and no such study had been performed on this strain of *Acaryochloris*.

Chapter 3 details the isolation and characterisation of two phages, A-HIS1 and A-HIS2, which were isolated on *A. marina* strain MBIC11017. This work was largely motivated by the observation that all known organisms are infected by viruses, therefore it was hypothesised that *Acaryochloris* spp. may be used as hosts to isolate lytic phages. *A. marina* also represents a new model for studying cyanophages as they are sedentary cyanobacteria as opposed to free-living in the marine environment.

In **Chapter 4**, a genomic and proteomic characterisation of these phages is presented. Genomic DNA was extracted from phages A-HIS1 and A-HIS2 and sequenced. Following this, phage genes were predicted and compared to database sequences to annotate the genomes. This was combined with mass spectrometry to identify phage structural genes. The sequence data was then used to construct phylogenetic trees to access the evolutionary history of these phages and to support the annotations and thus the putative origin of genes.

The acquisition of phage genome data allowed the design and implementation of experiments to assess the expression of phage genes during infection using qPCR. Results from these experiments are presented in **Chapter 5**, which provides an insight into phage-host interactions with regards to *A. marina* and phages A-HIS1 and A-HIS2. Phage adsorption experiments and the effect of phage infection on photosynthesis are also investigated in **Chapter 5**. These experiments were motivated by previous studies on the parallel cyanobacteria-phage system: *Synechococcus* sp. WH7803 and phage S-PM2 ((Clokie et al., 2006a; Shan et al., 2008)). Upon infection, *Synechococcus* sp. WH7803 has been shown to undergo significant changes in light-harvesting apparatus, which may also be true of *Acaryochloris* spp. after phage infection. The study of phages which infect *Acaryochloris* spp.

may yield new insights into phage-host dynamics.

In **Chapter 6**, biofilms of *A. marina* strain MBIC11017 were investigated. *Acaryochloris* spp. have been found in microbial biofilm communities together with other bacteria and it was therefore hypothesised that they may form biofilms by themselves. These biofilms could then be used to assess the metabolism of *Acaryochloris* cells grown in conditions close to their habitual nature, rather than in liquid media for example. Simultaneously, there was motivation to develop scanning electrochemical microscopy (the microscope was subsequently assembled and optimised in-house) to probe such biological systems using markers of photosynthesis such as oxygen evolution to determine the heterogeneity of cells.

Finally, in **Chapter 7** a conclusion is presented along with future directions the work can be taken in.

Chapter 2

Pigment adaptation and composition in *Acaryochloris* spp.

2.1 Introduction

Acaryochloris strains have been isolated from several varied locations and habitats worldwide demonstrating a diverse and dynamic ecology. Preliminary observations of strains MBIC11017 and CCMEE5410 (Miyashita et al., 1996; Miller et al., 2005), the symbiotic/epibiontic strain isolated from a colonial ascidian and the free-living epilithic strain, respectively, indicated that these two strains changed colour under different light conditions. In this chapter, absorbance and fluorescence spectroscopy, SDS-PAGE[1], western blots, mass spectrometry and LSCM (laser scanning confocal microscopy) were used to assess the pigment compositions of these two strains.

A. marina uses a combination of phycobiliprotein (PBP) rods and photosystem complexes to harvest light (§ 1.2.4). Although the presence of PBPs has been verified (Marquardt et al., 1997; Hu et al., 1998), little is known about their response to different light regimes. The pigment composition of *Acaryochloris* strain CCMEE5410 was characterised by Miller et al. and it

[1]sodium dodecyl sulphate polyacrylamide gel electrophoresis

was noted from electron micrographs that the thylakoid membranes lacked phycobilisomes (Miller et al., 2005). Previously, no study has focussed on the effect of light quantity on the pigment composition of *Acaryochloris* spp. and this work was published in 2007 (appended to thesis, (Chan et al., 2007)).

The subject of algal and cyanobacterial pigments and species distribution in the water column has been debated for many years starting with Engelmann putting forward his theory of chromatic adaptation in 1883 in which he stated that algal pigments are controlled by the quality of light at a particular depth (Engelmann, 1883). An alternative theory was proposed by Oltmanns in 1892 (Oltmanns, 1892), in which he proposed that it was the quantity of light and not the quality which controlled the depth distribution. It should be noted that the results of this chapter were obtained with non-axenic cultures, and therefore should be interpreted in this context. All experiments in other chapters were performed with axenic culture unless otherwise stated.

2.2 Materials and methods

All glassware, solutions, media or apparatus were autoclaved at 121 °C for 15 min using an ST3028 portable autoclave (Dixons Surgical Instruments Ltd.). For all experiments in this thesis, 18.2 MΩ ultrapure water (UPW) was used.

2.2.1 Media

Artificial seawater (ASW) is made up of two parts, which are prepared separately: the main salts listed in Table 2.1 and a trace metal solution (Table 2.2). The recipe is detailed in (Clokie and Kropinski, 2009; Millard, 2009) and is based on that of (Wyman et al., 1985). The main difference from the original recipe is the use of iron (III) chloride instead of ferric

ammonium citrate at half the concentration. The salts were made up at higher concentrations in UPW separately and then combined to make the 2× ASW salts solution which was adjusted to pH 8.0. This was then autoclaved and stored at room temperature. All compounds were dissolved separately in UPW except for $FeCl_3 \cdot 6H_2O$, which was dissolved in a small volume (3–5 mL) of concentrated HCl. The dissolved compounds were then combined and made up to 1 L with UPW and stored at 4 °C. Typically, a 2× salts solution was diluted to 1× and 1 mL of trace metal solution was added to make 1× ASW, which was then autoclaved and ready for use.

Table 2.1: Composition of salts in ASW.

Ingredient	Concentration (g L^{-1})
$NaNO_3$	0.75
$MgCl_2 \cdot 6H_2O$	2
KCl	0.5
$CaCl_2 \cdot H_2O$	0.5
$MgSO_4 \cdot 7H_2O$	3.5
Tris	1.1
K_2HPO_4	0.03
NaCl	25

Table 2.2: Composition of trace metal stock.

Ingredient	Concentration (g L^{-1})
H_3BO_3	2.86
$MnCl_2 \cdot 4H_2O$	1.81
$ZnSO_4 \cdot 7H_2O$	0.222
$Na_2MoO_4 \cdot 2H_2O$	0.39
$CuSO_4 \cdot 5H_2O$	0.008
$Co(NO_3)_2 \cdot 6H_2O$	0.049
$FeCl_3 \cdot 6H_2O$	3.0
EDTA (Na_2Mg)	0.5

2.2.2 Culture maintenance

Strains MBIC11017 and CCMEE5410 were kindly provided by Wolfgang Hess and Michelle Wood respectively. Strain stocks were sub-cultured about

once a month or every two months and maintained in conical flasks in autoclaved ASW. Cultures were kept at 28 °C on an illuminated shelf under 10 μmol photons $m^{-2}s^{-1}$ of continuous white fluorescent light (Osram Ltd. L65/80W/23).

2.2.3 Exponential phase culture

Cultures were grown in autoclaved ASW in 1 L polycarbonate NALGENE 4105 Fernbach flasks or 2 L glass conical flasks under 30–50 μmol photons $m^{-2}s^{-1}$ of continuous white fluorescent light. Cultures were aerated with Interpet AP4 aquarium air pumps and stirred using magnetic fleas.

2.2.4 Pigment adaptation growth conditions

Autoclaved ASW was inoculated with culture in the exponential phase in 50 mL BD Falcon polystyrene tissue culture flasks. Cultures were maintained at 23 °C or 28 °C under continuous white fluorescent light. The range of light intensities used were 10, 20, 30 and 40 μmol photons $m^{-2}s^{-1}$. Cultures were also grown at 28 °C under 25, 50 and 100 μmol photons $m^{-2}s^{-1}$ at 125 rpm on a New Brunswick Scientific Innova™ 2000 platform shaker. Light intensity was determined using a Skye Instruments Ltd. SKP 200 with lux sensor.

2.2.5 Cryogenic stocks

Once cells were made axenic, cryogenic stocks were created. Cells were grown to the exponential phase and concentrated by centrifugation. Following this, dimethyl sulfoxide (DMSO) was added to a final concentration of 7 % (v/v) and cells were frozen and stored in liquid nitrogen (Day, 2007).

2.2.6 Cleaning agar

The method for cleaning DIFCO Bacto Agar for use with cyanobacteria is detailed in (Clokie and Kropinski, 2009; Millard, 2009). Typically, 200 g of agar was mixed with 4 L of UPW. The agar was left to settle for about 30 min and then excess UPW was removed. The moist agar was then poured into a Buchner funnel lined with 3MM Whatman filter paper, which was attached to a vacuum pump via a series of borosilicate conical flasks with side-arms. Once the agar was sufficiently dry, the filtrate was removed and this process was repeated until the filtrate was clear. The agar was then washed once with ethanol followed by acetone. The agar was then spread out onto 3MM Whatman filter paper and air-dryed. The cleaned agar was then stored in a plastic container at room temperature.

2.2.7 Cyanobacterial culture plates

10 g of cleaned agar in 500 mL of UPW and 500 mL of 2× ASW were autoclaved separately and then combined. Once the molten agar had cooled to 50 °C it was poured into Petri dishes. Plates were briefly flamed after pouring to remove air bubbles and allowed to set. Plates were then stored at 4 °C. Plates were warmed to the required temperature before use.

2.2.8 Plate streaking and clonal culturing

To generate single colonies a sterile wire loop was used to streak out culture onto the appropriate plates. Single colonies were then either re-streaked or transferred to liquid media.

2.2.9 Contamination plates

For 40 plates, 500 mL of autoclaved 2× ASW was added to an autoclaved 500 mL mixture containing final concentrations of 2 % (w/v) glucose, 0.15 % (w/v) yeast extract and 1.5% (w/v) DIFCO Bacto Agar and UPW. Plates

were poured and stored as for § 2.2.7. To test for contamination, 0.1 mL of culture was streaked onto a plate and the plates were wrapped in foil and incubated at 28 °C for at least a week. Alternatively, (Luria-Bertani) LB-agar plates were used (10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl and 15 g bacto-agar in 1 L of UPW) (Maniatis et al., 1982b).

2.2.10 Flow cytometry

Cells were counted on a Becton Dickinson FACScan containing a 488 nm Argon laser. In the FACScan scattered light is detected either as forward scatter (FSC), which is light deflected less than 10° after interaction with the sample or side scatter (SSC), which is light deflected 90° to the laser orientation after interaction with the sample.

Table 2.3: FACScan channels and filters.

Parameter/Channel	Filter
FSC	488/10
SSC	
FL1	530/30
FL2	585/42
FL3	650LP

FSC is used to infer particle size and SSC is used to infer particle granularity, which gives an idea of the content of the particle by restricting or allowing light to pass through. Green, orange and red fluorescence of a particle are also recorded in the channels FL1, FL2 and FL3, respectively. The five parameters detected and the corresponding filters are summarised in Table 2.3. UPW was used as sheath fluid. Samples were diluted with autoclaved ASW before acquisition. The cells mL^{-1}, x, for a batch was calculated using Equation 2.1 taking into account the dilution factor, d, flow rate, f (mL min^{-1}) the number of cells, C, counted counted by the flow cytometer for each 1 mL diluted sample in 30 s.

$$x = \frac{2Cd}{f} \tag{2.1}$$

2.2.11 Absorbance measurements

Optical density or absorbance single wavelength measurements of *Acaryochloris* strains were measured on a Spectronic Instruments 8100 Series spectrofluorometer at room temperature in 1 mL plastic cuvettes. Samples grown under static conditions were shaken before sampling. Absorbance spectra were recorded at room temperature on a UV 500 UV/VIS spectrometer using Vision32 software with ASW as the reference sample. To adjust spectra for cell light scattering, they were normalised to cell light scattering data, which was generated from absorbance data from boiled cells. Samples grown under static conditions were shaken before sampling. ASW was used as the reference sample. The absorbance index of Chl d, $\mathrm{I}_{\mathrm{Chl}d}$, was calculated using

$$\mathrm{I}_{\mathrm{Chl}d} = \frac{A_{716\ \mathrm{nm}} - A_{800\ \mathrm{nm}}}{A_{800\ \mathrm{nm}}} \tag{2.2}$$

to give a measure of Chl d per cell, where $A_{x\ \mathrm{nm}}$ is the absorbance at x nm based on similar equations used by Six et al. for *Synechococcus* sp. WH8102 (Six et al., 2004).

2.2.12 Fluorescence measurements

Fluorescence spectra were recorded at room temperature on a Spectronic Instruments 8100 Series spectrofluorometer. Fluorescence spectra were smoothed using the Savitzky-Golay algorithm (Savitzky and Golay, 1964). Samples grown under static conditions were shaken before sampling.

2.2.13 Protein extraction

Cells were harvested and pelleted, flash frozen in liquid nitrogen and stored at -20 °C. Thawed cells were washed and resuspended in 10 mM HEPES[2], pH 7.4, bead beaten using 0.17–0.18 mm Glasperlen (B. Braun Biotech International GmbH) for 15 min min at room temperature and the crude cell extract was removed after centrifugation at 16,000 g.

2.2.14 Protein concentration determination

The protein concentration of each sample was assessed using a BCA (bicinchoninic acid) Protein Assay Kit in 96 well plates (Becton Dickinson) and a Labsystems iEMS Reader MF with Ascent Software Version 2.4.2.

2.2.15 Protein gel electrophoresis

A BioRad Mini-Gel kit was used to cast polyacrylamide gels with a 12 or 15 % resolving gel and a 5 % stacking gel. For example, 50 mL of 12 % resolving gel solution was obtained by combining 15 mL of 40 % (w/v) acrylamide solution (Fisher BioReagents), 12.5 mL 1.5 M Tris-HCl (pH 8.8), 0.5 mL 10 % (w/v) SDS and UPW to 50 mL. Similarly, to make 25 mL of stacking gel solution, 3.13 mL of 40 % (w/v) acrylamide solution was added to 6.3 mL 0.5 M Tris-HCl (pH 6.8), 0.25 mL of 10 % (w/v) SDS and UPW to 25 mL. For one gel, 50 μL of ammonium persulfate (APS, Fisher Scientific) and 10 μL of tetramethylethylenediamine (TEMED, Fluka) were added to set 5 and 2 mL of the resolving and stacking solutions, respectively. First, the resolving gel was allowed to set in the gel-caster (placed at 28 °C to set faster). Ethanol was added on top of the resolving gel layer remove air bubbles. The ethanol was subsequently removed after the resolving gel was set, and the stacking gel solution was poured on top, followed by the insertion of the well-comb into the gel-caster. Samples were prepared after thawing by

[2] 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

adding a 5× denaturing mix (5 % (w/v) SDS, 27 % (v/v) glycerol, 98 mM Tris-HCl pH 6.8, 1.3 % (v/v) β-mercaptoethanol and 0.2 % (w/v) bromophenol blue) and heating at 100 °C for 3–5 min. The gels were run at 200 V for $\sim$1 h. After electrophoresis, gels were stained with Coomassie-blue in 3 % (v/v) acetic acid and destained in water.

2.2.16 Mass spectrometry

Proteins from polyacrylamide gels were identified by LC-MS/MS[3]. Individual bands were excised and treated according to the protocol of Aitken and Learmonth (Aitken and Learmonth, 2002). Briefly, gel pieces were incubated with three changes of 200 mM ammonium bicarbonate (ABC)/50 % (v/v) acetonitrile (ACN) at 30 °C for 30 min min to remove SDS. The proteins were reduced by incubation with 20 mM DTT (dithiothreitol)/0.2 M ABC/50 % ACN at 30 °C for 1 min. After several washes to remove DTT, the cysteine residues were alkylated by incubation with 50 mM iodoacetamide in fresh 200 mM ABC/50 % ACN for 20 min in the dark. After further washing in 20 mM ABC/50 % ACN, the band was cut into small (1×2 mm) pieces and shrunk with the addition of 100 % ACN. The gel pieces were then rehydrated with 0.5 μg modified sequencing grade trypsin (Promega), made up in 50 mM ABC (pH 7.8), on ice for 15 min. 100 μL of 50 mM ABC, pH 7.8 was added to cover the pieces and incubated overnight at 30 °C. The peptides contained within the supernatant were transferred to another tube and further extraction was achieved by the addition of another 100 μL 200 mM ABC/50 % ACN and sonication for 30 min at 35 °C. The supernatants were collected, combined and lyophilised in a centrifugal evaporator for 30 min. The aqueous solution containing digested peptides was passed through a 0.22 μm filter (Millipore, MA), then loaded onto a 15 cm, 75 μm internal diameter, C18 media column (LC packings, Dionex,

[3]liquid chromotography-mass spectrometry/mass spectrometry

CA) and eluted with a 5–95 % gradient of water:acetonitrile (0.1 % formic acid) over 1 h at a flow rate of 0.2 μL min^{-1} directly into a Thermofinnigan LCQ DECA quadropole ion trap instrument using online nanoHPLC-MS/MS. MS/MS data was collected using data dependent acquisition experiments on the major ion species using the accompanying Excalibur software (Thermo Electron Corp., San Jose, CA). The raw data from the LCQ was processed using SEQUEST (Eng et al., 1994) to establish parent:daughter ion relationships, producing a series of .dta files for each LC-MS/MS run. The merged .dta files were then submitted to the MASCOT server[4] for searching against the entire NCBI[5] nr (non-redundant) database.

2.2.17 Western blots

Samples for Western blotting were analysed on 16 % SDS-PAGE gels. To increase the resolution for small proteins 1.25 % (w/v) solid Bis-Tris was added and the concentration of APS and TEMED was doubled to 0.2 % (w/v) and 0.005 % (w/v), respectively (compared to § 2.2.15). Gels were blotted onto Hybond P (Amersham Biosciences) using the Towbin-buffer system. The membrane was blocked in phosphate buffered saline with 0.1 % Tween (PBS-T) with 3% skimmed milk powder for 1 h followed by three 30 min washes in PBS-T. The primary antibody (rabbit anti-APC polyclonal antibody, unconjugated from AbD Serotec (Biogenesis) obtained from being challenged with APC and PC in *Spirulina platensis*) was incubated for 1 h at 1:1000 and the anti-rabbit IgG HRP conjugate antibody (Promega) was incubated at 1:10000 for 1 h. This was washed with PBS-T and then the detection reagents were added for 1 min (Amersham, ECL-kit). The blots were developed using an AGFA Curix 60 on Fuji Medical X-ray film for 5–30 s.

[4] `http://www.matrixscience.com/`
[5] National Center for Biotechnology Information.

2.2.18 Chromophore detection

Bilin-containing bands were observed under UV and imaged in GeneFlash on a Syngene Bio Imaging transilluminator unit. Visualisation of bilin-containing bands under UV was enhanced by incubating the gels in 2 M glycine with 0.2 M $ZnSO_4$ for 30 min (Berkelman and Lagarias, 1986).

2.2.19 Laser scanning confocal microscopy

Liquid cell culture samples were spotted onto 1 % agar-ASW plates and left to dry on the bench for 10–20 min. LSCM was performed using a Leica SP2 and a Leica DM RE7 upright microscope. Blocks of agar with adsorbed cell samples were cut from the agar plate and placed into the well of a custom built sample holder. A coverslip was placed flush over the sample and then imaged using a 63$\times$ oil immersion lens. PBP and Chl *d* fluorescence were detected between 640–670 nm and 700–750 nm respectively by excitation with a 633 nm red laser. The same gains were used for the PBP (400$\pm$1 V) and Chl *d* (490$\pm$1 V) channels to compare shaken MBIC11017 cultures grown at 25, 50 and 100 μmol photons $m^{-2}s^{-1}$. The offset in both channels was set to 2$\pm$1 %. Where comparison between images from the same channel was made, the same gains and offset conditions were used. In Volocity (Version 3.7.0) from Improvision® the maximum intensity projection (MIP) contribution was adjusted for all images from 100 % (acquired image) to 200 % to give maximum contrast.

2.3 Results and Discussion

2.3.1 Pigment composition of *Acaryochloris* spp.

Strain MBIC11017 has been cultured under a range of conditions in several laboratories, at light intensities ranging from 5–80 μmol photons $m^{-2}s^{-1}$ and either under continuous light or light/dark regimes. Temperatures have

ranged from 20 °C to 30 °C. Aeration and in particular shaking and stirring of cultures have been sporadically used. Strain CCMEE5410 has only been cultured under 15 and 50 μmol photons $m^{-2}s^{-1}$ of continuous fluorescent light (Miller et al., 2005). Initial observations during culturing of strains MBIC11017 and CCMEE5410 indicated a significant difference in the pigmentation of these two strains of *A. marina*. Under a range of temperatures and light intensities, the CCMEE5410 cultures were yellow, becoming yellow-green as cultures aged. In contrast, the MBIC11017 cultures were bright lime green, turning a dark green, the more dense the culture. However, under certain conditions, MBIC11017 cultures became yellow and virtually indistinguishable from the CCMEE5410 cultures. The whole cell absorbance for strains MBIC11017 and CCMEE5410 are shown in Fig. 2.1. The spectra for strain MBIC11017, grown at 28 °C, under 10–20 μmol photons $m^{-2}s^{-1}$ of continuous white fluorescent light were consistent with previously published spectra (Miyashita et al., 1996, 1997; Mimuro et al., 2000).

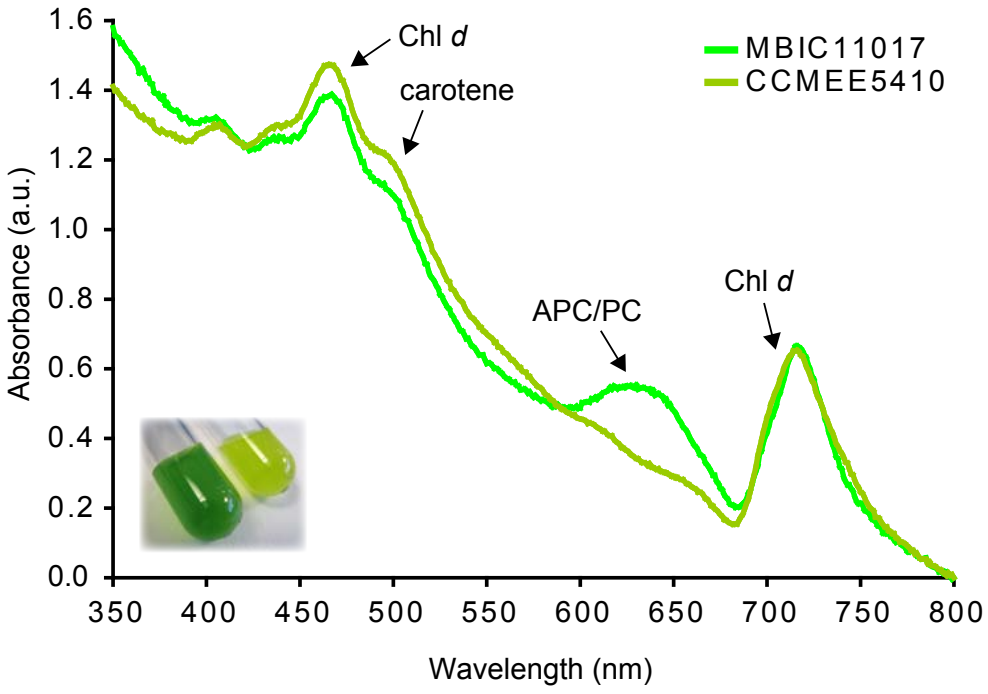


Fig. 2.1: *Acaryochloris* spp. spectra. Absorbance spectra of strains MBIC11017 and CCMEE5410, normalised at 800 nm.

Compared to strain MBIC11017, strain CCMEE5410 exhibited no phycobilin signal. Furthermore, spectra for strain CCMEE5410 did not demonstrate a significant peak at 680 nm where one would expect the Chl *a* Q_y absorption peak, although this is not surprising given the low level of Chl *a* which has been reported for *A. marina* (Miller et al., 2005). In light of this, we assigned wavelengths between 590 and 680 nm from the whole cell spectra to be due to phycobilin absorption compared to Schiller et al. where previously the phycobilin band was assigned between 580 and 660 nm (Schiller et al., 1997). The Chl *d* Q_y absorption peak lay between 715–717 nm after light scattering correction for both strains.

2.3.2 Pigment adaptation of *Acaryochloris* spp.

2.3.2.1 Absorbance spectra

To determine how light intensity and temperature affected pigment composition, cultures were grown under varying light intensities and at two different temperatures. Strains MBIC11017 and CCMEE5410 both demonstrated a pigmentation adaptation mechanism. Cultures began to adapt after 2 days and the adaptation was prominent after 4 days. To compare spectra, the absorbance data were normalised at 580 nm, which revealed the relative changes in the most prominent pigmentation peaks in whole cell spectra: those of Chl *d* and the PBPs. In MBIC11017 cultures grown at 28 °C, relative absorbance spectra showed that an increase in light intensity lowered the absorbance of both Chl *d* and the phycobilins of intact cells (Fig. 2.2(A,B)). Similarly, relative absorbance spectra in strain CCMEE5410 showed a gradual decrease in the Chl *d* peak at 716 nm with respect to increasing light intensity Fig. 2.2(C,D).

Similar results were obtained at 23 °C as at 28 °C. To compare the effect of temperature, the Chl *d* absorbance index, I_{Chld}, was calculated for cultures grown at 23 °C and 28 °C for strains MBIC11017 and CCMEE5410.

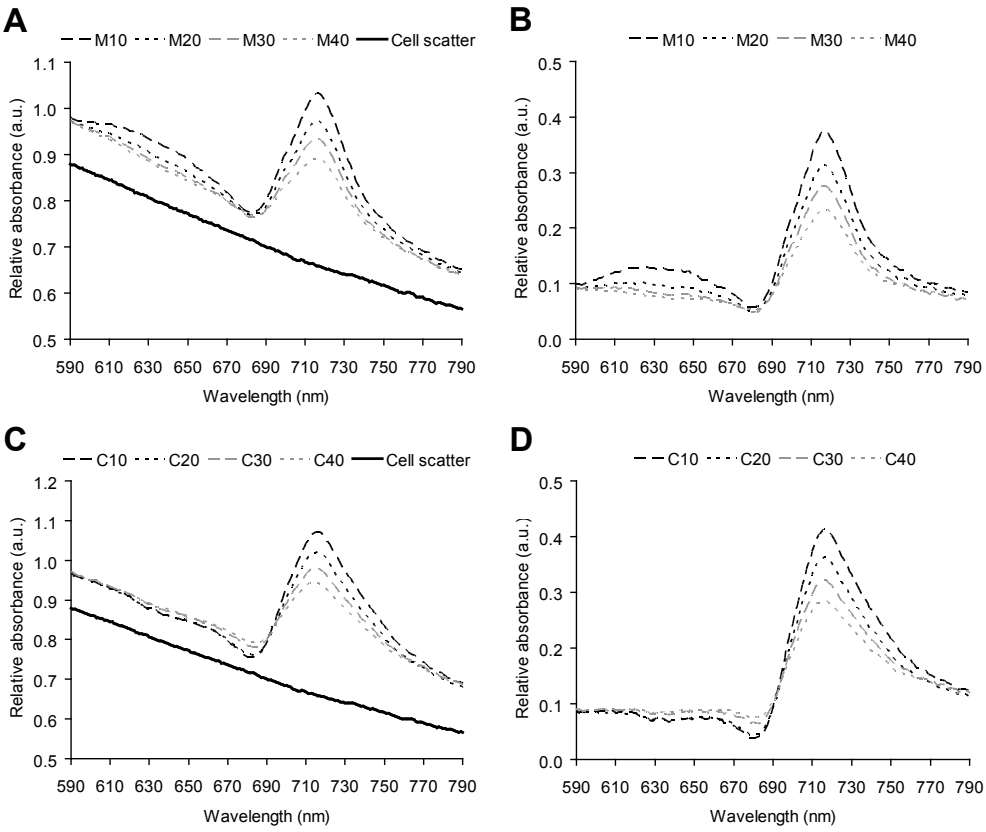


Fig. 2.2: Effect of irradiance on absorbance spectra. Absorbance spectra of strains MBIC11017 (M) and CCMEE5410 (C) grown at 28 °C under continuous white light at 10, 20, 30 and 40 μmol photons $m^{-2}s^{-1}$ 10 days after inoculation after light adaptation. Whole cell spectra were normalised at 580 nm (A and C) and the corresponding light scattering corrected spectra were obtained after the subtraction of cell scatter data (B and D). The figure legends indicate strain and light intensity.

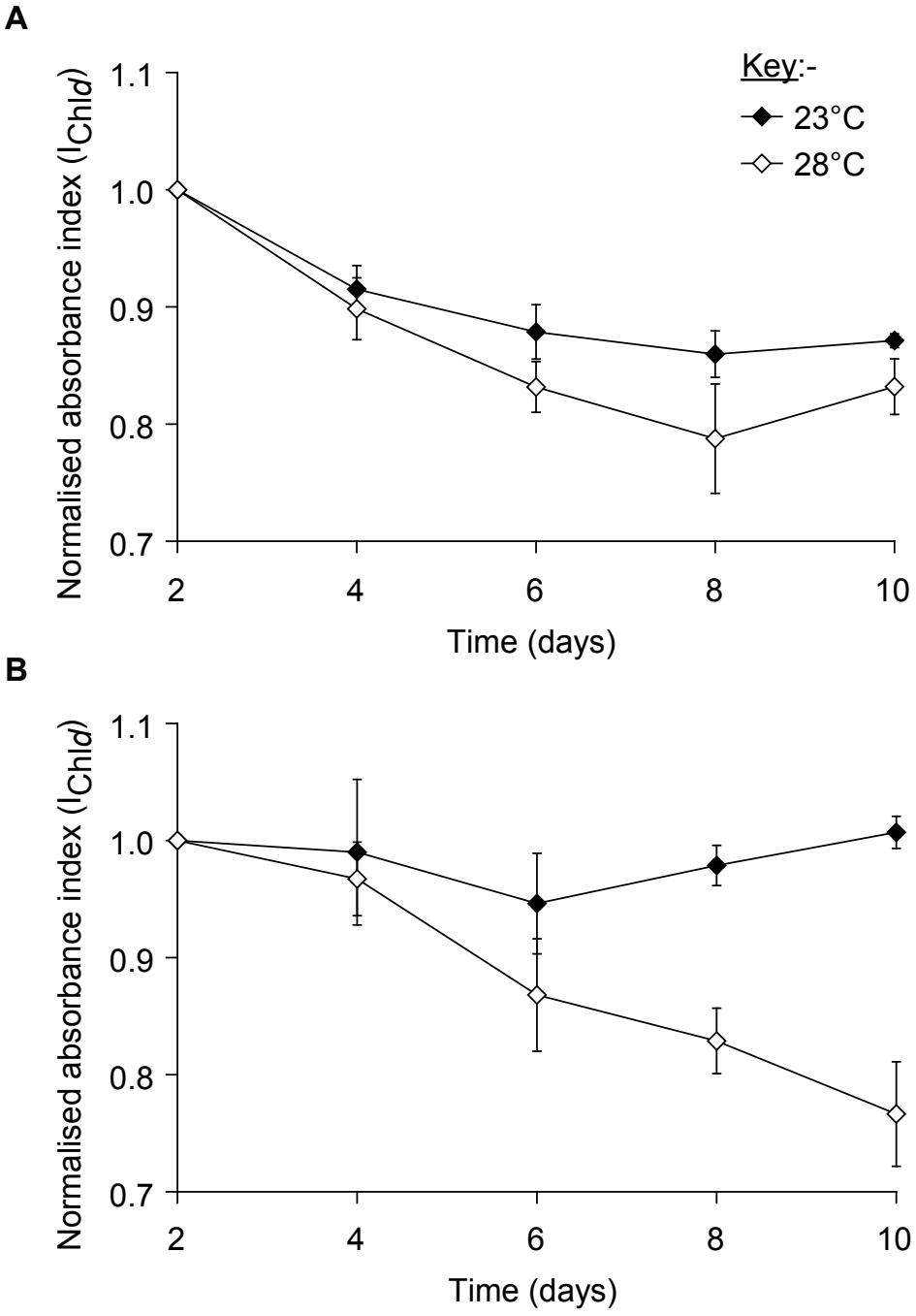


Fig. 2.3: Effect of temperature on pigment adaptation. The absorbance index of Chl *d*, I_{Chld} (see § 2.2.11), was normalised to the first time point for strains (A) MBIC11017 and (B) CCMEE5410 grown at 23 °C and (B) 28 °C at 40 μmol photons $m^{-2}s^{-1}$. Error bars are one standard deviation, n=3.

$I_{Chl\,d}$ is a measure of the Chl *d* per cell (§ 2.2.11, (Six et al., 2004)), which showed that cultures were slower to adapt at 23 °C than at 28 °C for both strains (Fig. 2.3). Previously, it has been shown that the optimum growth temperature for *A. marina* is 28 °C and therefore the processes for pigment adaptation is likely to be optimised at this temperature (Miyashita et al., 2003).

2.3.2.2 Fluorescence measurements

Emission and excitation spectra for both strains were recorded after pigment adaptation to investigate whether the fluorescence dynamics of the whole cells were affected by growth at different light intensities. Previously, Petrášek et al. showed that the intensity of PBP fluorescence is highly dependent on the excitation wavelength (Petrášek et al., 2005). This was observed during preliminary optimisations. In light of this, emission spectra were measured by exciting at 630 nm to demonstrate relative fluorescence intensities of PBPs and Chl *d* in the two strains. Strain MBIC11017 showed characteristic peaks around 650 nm and 717 nm in emission spectra (Fig. 2.4(A)), which are attributed to PBPs and Chl *d*, respectively. A gradual decrease in these peak wavelength bands was observed with respect to increasing growth light intensity, indicating the loss of PBPs and Chl *d* at higher light intensities. Strain CCMEE5410 showed a similar decrease in the Chl *d* peak at 725 nm with respect to an increase in growth light intensity, and again no peak corresponding to PBPs was observed.

Excitation spectra were recorded by collecting the emission at 717 nm and 725 nm for strains MBIC11017 and CCMEE5410, respectively Fig. 2.4(B), i.e. the emission wavelengths assigned to Chl *d* for each strain. The excitation spectra were measured between 510–690 nm to show the degree of excitation energy transfer from PBPs to Chl *d* in strain MBIC11017,

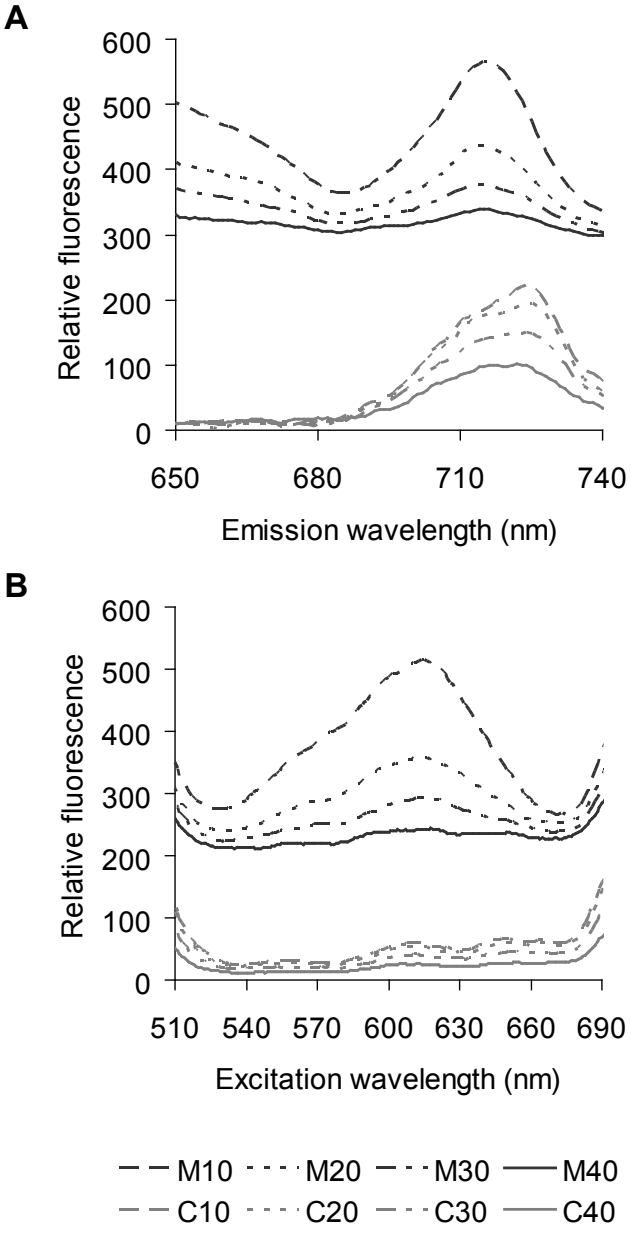


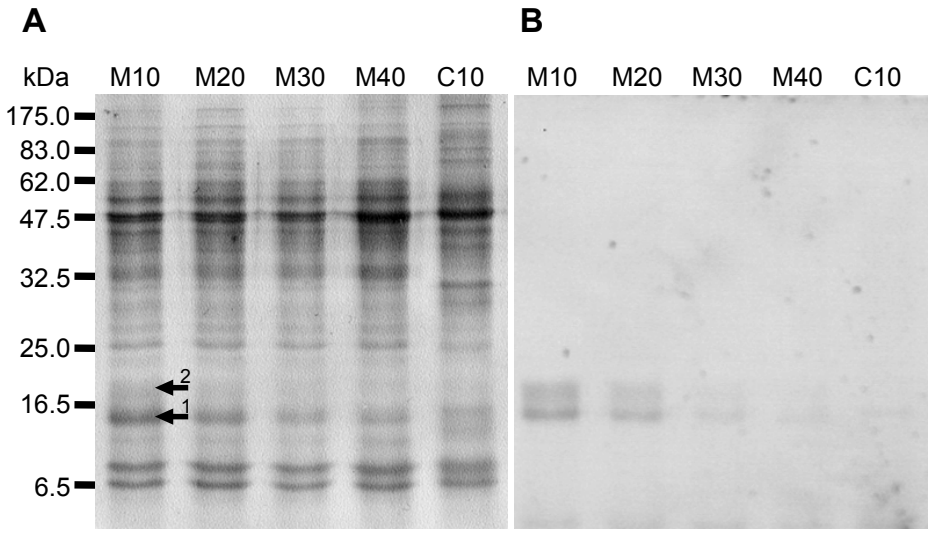
Fig. 2.4: Effect of irradiance on fluorescence spectra. (A) Emission and (B) excitation spectra of strains MBIC11017 and CCMEE5410 grown under different light regimes. Emission spectra were obtained by excitation at 630 nm for both strains, while for excitation spectra the emission wavelengths at 717 nm and 725 nm were used for strains MBIC11017 (M) and CCMEE5410 (C), respectively. Both strains were grown at 28 °C under continuous white light at 10, 20, 30 and 40 μmol photons $m^{-2}s^{-1}$. Spectra were recorded at room temperature 10 days after inoculation after light adaptation and normalised relative to the $OD_{800\,nm}$ of the same samples. The figure legend indicates strain and light intensity.

which demonstrated a decrease in signal intensity at higher growth irradiances. For strain CCMEE5410 the excitation spectra of all samples resembled that for strain MBIC11017 at 40 μmol photons $m^{-2}s^{-1}$, though there was a gradually decreased signal over the growth irradiance range used. The emission spectra of strain CCMEE5410 indicates that either Chl *d* or some other pigment is able to absorb light at 630 nm and contribute to Chl *d* fluorescence emission in the absence of PBPs Fig. 2.4(A). Presumably, MBIC11017 cells exposed to higher light intensities adapt by losing their PBPs as they no longer require them to increase the efficiency of light-harvesting (Boichenko et al., 2000; Petrášek et al., 2005), which is consistent with the predominantly low light, yet fluctuating environment in which *A. marina* is found.

2.3.2.3 Protein analysis

SDS-PAGE of soluble protein from whole cell samples (grown at 28 °C under continuous white light at 10–40 μmol photons $m^{-2}s^{-1}$, revealed a decrease in the intensity of the chromophore bands, visible under UV, corresponding to PC and APC (Marquardt et al., 1997) Fig. 2.5(B) for strain MBIC11017 with respect to increasing growth irradiance. The crude soluble protein from strain CCMEE5410 cells grown under the same conditions appeared to completely lack chromophore bands corresponding to PC and APC under UV illumination (Fig. 2.5(A, B), lane C10). Due to the high degree of similarity between the lanes (corresponding to different growth light intensities) for strain CCMEE5410, only the lane for cells grown at 10 μmol photons $m^{-2}s^{-1}$ is shown (lane C10).

LC-MS/MS analysis of each extracted protein band and subsequent MASCOT searches identified the proteins present in the chromophore bands for strain MBIC11017 as being the α and β subunits of phycocyanin Fig. 2.5(C). Phycocyanin peptides identified here match those found in several cyano-



C Peptides identified by LC-MS/MS

(1) C-phycocyanin 2 α chain (*Calothrix* sp., accession number: B28539)

```
  1 MKTPLTEAVA TADSQGRFLS STELQVAFGR FRQASASLDA AKALSSKANS
 51 LAQGAVNAVY QKFPYTTQMQ GKNFASDQRG KDKCARDIGY YIRIVTYCLV
101 AGGTGPLDDY LIGGLAEINR TFDLSPSWYV EALKYIKANH GLSGDPAVEA
151 NSYIDYAINA LS
```

(2) Phycocyanin β subunit (*A. marina*, accession number: Q5GJ62)

```
  1 KLLSVVSDGN KRMDVVNRIT SNASTIVANA ARDLFEEQPS LIQPGGNAYT
 51 HRRMAACLRD MEIILRYVTY AIFAGDASIL EDRCLNGLKQ TYQTLGVPTK
101 SMSLSVSKMR DAALEIASDP NGVTQGDCSS LISEVSDYFD LAARAVG
```

D

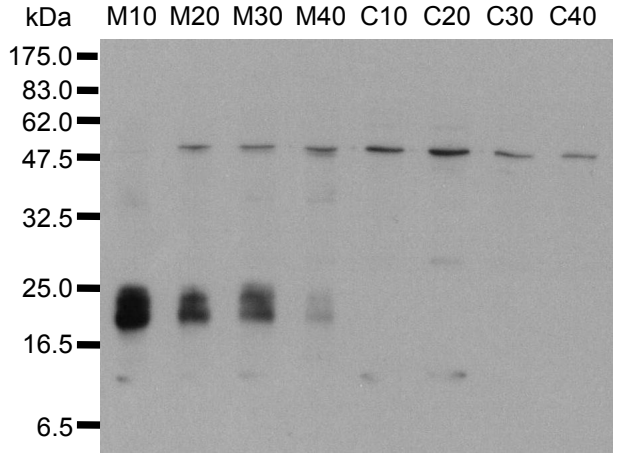


Fig. 2.5: PBP protein and chromophore characterisation. (A) Coomassie-stained SDS-PAGE gel, (B) UV image of soluble protein extracts from strains MBIC11017 CCMEE5410. Cells were grown at 28 °C under continuous white light at 10, 20, 30 and 40 μmol photons $m^{-2}s^{-1}$. Peptides identified from LC-MS/MS are highlighted in bold for the bands indicated in the SDS-PAGE gel with arrowheads. (D) Western blot of soluble protein extracts from strains MBIC11017 and CCMEE5410. Cells were grown at 28 °C under continuous white light at 10, 20, 30 and 40 μmol photons $m^{-2}s^{-1}$. Lanes are labelled by strain (M=MBIC11017 and C=CCMEE5410) and light intensity (western blot contributed to this work by Anja Nenninger and mass spectrometry was performed by Samuel Clokie).

bacteria. Using the *A. marina* MBIC11017 genome data[6] ll peptides were identified which were predicted in the *A. marina* genome. Bands were removed from the corresponding 10–20 kDa size region in strain CCMEE5410 and analysed using LC-MS/MS did not reveal any evidence of peptides corresponding to PC or APC proteins. Indeed, the peptides which were recovered matched a DNA-binding stress protein (*Synechococcus* sp. strain PCC6301, AN (accession number): Q5N285_SYNP6), thioredoxin (*Synechococcus* sp. strain WH8102, AN: Q7U898_SYNPX) and photosystem I reaction centre subunit II (photosystem I 16 kDa polypeptide; PSI-D, *Synechococcus elongatus*, AN: PSAD_SYNEL).

This result was further strengthened by western blotting (Fig. 2.5) which shows that in MBIC11017 the amount of APC and PC is reduced under higher light conditions and they are absent in CCMEE5410. The western blot showed clear bands of ∼58 kDa in all samples (it is very faint in lane M10, but present). These bands are thought to be due to the non-specific binding of the antibody to an unknown protein, since no corresponding chromophore band was visible under UV.

2.3.2.4 Laser scanning confocal microscopy

LSCM was an extremely useful method to examine the pigment composition of *A. marina* grown under different light regimes. Confocal images showed that CCMEE5410 cells had no signal in the PBP channel Fig. 2.6(A). In comparison, for MBIC11017 cells grown at 100 μmol photons $m^{-2}s^{-1}$ even a low PBP content is detectable Fig. 2.6(J), suggesting the complete absence of PBPs in strain CCMEE5410. Comparison of Fig. 2.6(A) and Fig. 2.6(B) shows that fluorescence from Chl *d* does not contribute to the PBP channel signal.

At the lower light intensity of 25 μmol photons $m^{-2}s^{-1}$, shaken MBIC11017

[6] http://genomes.tgen.org/

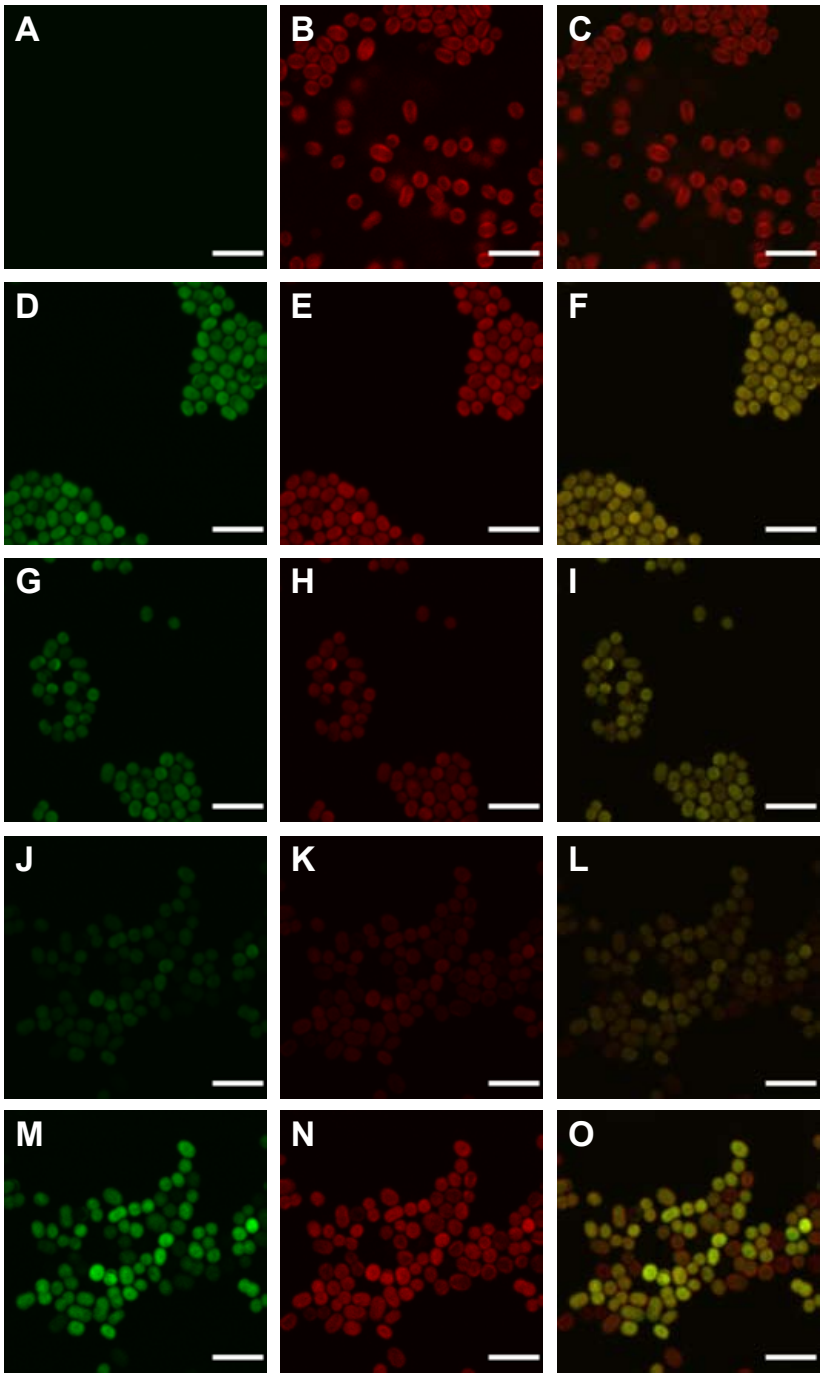


Fig. 2.6: *Acaryochloris* spp. confocal images. LSCM images of *A. marina* strains MBIC11017 and CCMEE5410. (A, B, C) CCMEE5410 cells. MBIC11017 cells grown at (D, E, F) 25, (G, H, I) 50 and (J, K, L) 100 μmol photons $m^{-2}s^{-1}$. Images M, N, O show the same cells as in J, K, and L but with increased gains in the PBP and Chl *d* channels. MBIC11017 cells were grown at 28 °C, whilst being shaken and harvested for imaging 5 days after inoculation. Images in green PBP channel (A, D, G, J, M) and red Chl *d* channel (B, E, H, K, N) with composite images (C, F, I, L, O), respectively. Scalebar, 6.3 μm.

cells showed strong signals in the PBP (Fig. 2.6(D)) and Chl *d* (Fig. 2.6(E)) channels. The decrease in fluorescence and abundance of PBPs and Chl *d* correspond to the absorbance and fluorescence data recorded for these cultures as observed for the non-shaken cultures (Fig. 2.2, Fig. 2.4). Under increased gains, shaken MBIC11017 cells grown at 100 μmol photons $m^{-2}s^{-1}$, contained a third subpopulation (Fig. 2.6(O)). This heterogeneous pigment composition is most likely attributed to natural shading, since it was observed in both shaken and non-shaken cells grown under higher light. The phenomenon of excitation energy transfer from PBPs to Chl *d* was particularly evident in these cultures.

In contrast, strain CCMEE5410 lacks PBPs, which implies that PBPs are not required for survival when *Acaryochloris* grow epilithically, in the shallow water and high light conditions of the Salton Sea, which is a highly eutrophic artificial lake maintained by agricultural runoff. Originally, strain CCMEE5410 was isolated from an epilithic mat and cultured under high light conditions of 50 μmol photons $m^{-2}s^{-1}$, presumably reflecting the environmental conditions in which it was discovered (Miller et al., 2005). The loss of PBPs in strain MBIC11017 at light intensities greater than or equal to 40 μmol photons $m^{-2}s^{-1}$ suggests it is likely that strain CCMEE5410 is a derivative of a PBP-producing strain akin to strain MBIC11017, which may have been inadvertently introduced into the Salton Sea from marine fish stocks (Ritchie, 2006). Prolonged exposure to high light intensities may have caused the ancestral strain to lose the ability to produce PBPs.

Comparison of the MBIC11017 and CCMEE5410 genomes should indicate whether the genes encoding PBPs are still present in CCMEE5410 or whether they are lacking altogether. Genomic data has since become available for strain MBIC11017 (Swingley et al., 2008), indicating the presence of genes involved in PBP biosynthesis and assembly. Additionally, no genes for phycoerythrin were found, although the presence of a phycoerythrin or

phycoerythrocyanin-like biliprotein was previously speculated (Hu et al., 1999). Duxbury et al. have since studied the effect of chromatic adaptation in strains MBIC11017 and CCMEE5410 and have shown that strain MBIC11017 chromatically adapts, while strain CCMEE5410 does so to a lesser extent due to the apparent absence of PBPs (Duxbury et al., 2009).

2.4 Conclusions

This is the first study to focus on the effects of growth irradiance on *Acaryochloris* spp. pigment composition. *Acaryochloris* strains MBIC11017 and CCMEE5410 were shown to adapt their pigment content under different light regimes using a number of techniques and cultures were slower to adapt at a lower temperature. Absorbance spectra showed that strain MBIC11017 appeared to lose its PBPs peak when grown at light intensities of 40 μmol photons $m^{-2}s^{-1}$ without shaking and 100 μmol photons $m^{-2}s^{-1}$ with shaking, although LSCM revealed that PBPs were still present in these cells albeit at levels not detectable by absorbance measurements. It was also concluded that the reduction in PBPs at higher light intensities corresponded to a reduction of both the protein and pigment components of the PBPs. *A. marina* is found in a dynamic environment with limited light and these properties are likely to be advantageous to the survival of *A. marina*.

Importantly, all results suggested that the free-living strain CCMEE5410 lacks PBPs - the antennae used to increase the efficiency of light-harvesting, which presumably is the result of an adaptation which led to its survival in the exposed epilithic habitat from which it was isolated (Miller et al., 2005). PBPs are therefore not necessarily required for the survival of *Acaryochloris* spp. strain CCMEE5410 may be a derivative of a PBP-producing strain akin to strain MBIC11017, which was introduced to the Salton Sea habitat (Ritchie, 2006).

Chapter 3

Isolation and characterisation of *Acaryochloris* phages

3.1 Introduction

The sheer abundance of bacteriophages and their inextricable relationship with bacteria has put them at the forefront of marine research (Torrella and Morita, 1979; Bergh et al., 1989; Suttle, 2005, 2007). Phages play key environmental roles in biogeochemical cycling, population genetics and bacterial evolution through horizontal gene transfer (Suttle, 2005). The study of phages which infect *Acaryochloris* spp. may yield new insights into phage-host dynamics and the effect of phages in biofilm-forming bacteria. Bacteriophages have never been isolated on *Acaryochloris* spp. or on a cyanobacterium with a sedentary lifestyle which has been shown to live either symbiotically with or epiphytically on a metazoan (Miyashita et al., 1996; Larkum and Kühl, 2005). This chapter details the isolation and basic characterisation of phages A-HIS1 and A-HIS2 which infect *Acaryochloris marina*, strain MBIC11017.

3.2 Materials and methods

3.2.1 Seawater samples

Seawater samples were collected in January 2006 by Remus Mohr from Heron Island (Kühl et al., 2005), Great Barrier Reef, Australia (-23° 26' 4.6032" S and 151° 55' 20.535" E) (Fig. 3.1). All seawater samples were stored in 50 mL falcon tubes and filtering was performed using 0.22 μm pore size filters (Millipore, MA). Filtered samples were transferred to new 50 mL falcon tubes, wrapped with aluminium foil and left at 4 °C before transportation.

Fig. 3.1: Aerial map of seawater sampling sites. *Acaryochloris* habitat, Heron Island, Great Barrier Reef, Australia. Reproduced from pers. comm. with R. Mohr, original image from Google Maps (`http://maps.google.com/`).

The sampling sites included the beach and the inner and outer reef edges. Sand samples were collected from the wet and dry sand zones and deposited in 50 mL falcon tubes until they were half-filled. The sand was then immersed in SM buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl and

8 mM $MgSO_4$) which had been kept at 4 °C. The tubes were wrapped with aluminium foil and then incubated on a turning wheel for 2 h. The supernatant was then filtered and stored at 4 °C in the dark.
Other samples were obtained by scraping the underside of the sea-squirt species *Lissoclinum patella* under which the *Acaryochloris* species have been found to grow (Miyashita et al., 1996). *Lissoclinum patella* was found growing on coral heads inside the reef flat and an additional water sample was obtained by incubating an ascidian itself in SM buffer. One half of the water sample was filtered while the other half was stored unfiltered.

3.2.2 Well assays

Exponentially growing strain MBIC11017 culture (§ 2.2.3) was aliquoted into 24 well plates. Following this, 50 μL of a seawater sample or 5–10 μL of phage stock (e.g. stocks as listed in Table 3.1) were added to a well containing exponentially grown culture. Plates were sealed with Parafilm M^{TM} and left at 23 °C or 28 °C. On the clearing of a well, the contents of the well were removed to a tube and cell debris was removed by centrifugation. Potential phage lysates were then stored at 4 °C in the dark. The same was performed for other cyanobacterial strains, however, these were grown under static conditions (§ 2.2.2, § 3.2.5).

3.2.3 Plaque assays

Plaque assays were performed by adapting the method of isolating phages described for *Synechococcus* WH7803 (Clokie and Kropinski, 2009; Millard, 2009). In short, MBIC11017 cells were grown to the exponential phase from an initial optical density (OD) of 0.01–0.05 and harvested for plaque assays at an OD (800 nm) between 0.4 and 0.5. Typically, 30–40 plaque assays were produced from 1 L of exponentially growing cells. Cells were centrifuged and the pellet was resuspended in 1$\times$ ASW to allow for 0.5 mL

of concentrated cells per plaque assay. Each 0.5 mL of cells was incubated for 1 h with 50 μL of seawater sample, mixed with 2.5 mL of 0.4 % (w/v) agar-ASW and then plated out onto 1 % (w/v) agar-ASW plates. A negative control was performed by adding 50 μL of 1$\times$ ASW instead of a seawater sample or phages sample. Plaque assays were incubated at 23 °C or 28 °C under 10–15 μmol photons $m^{-2}s^{-1}$ of white fluorescent light.

Once clear plaques had appeared, agar plugs in the clearings were removed from the Petri dish using a sterile Pasteur pipette. Plugs were transferred to 1 mL of ASW. Phage isolates were then made clonal by taking them through three rounds of plaque assays from single plaques. ASW lysates of clonal phages were stored at 4 °C in the dark after bacterial debris was removed by centrifugation.

3.2.4 *Synechococcus* phages

The cyanophage strains listed in (Table 3.1) were kindly provided by Andrew Millard and incubated with the desired bacterial strain in well assays (see § 3.2.2).

Table 3.1: Phage strains.

Phage	Reference
S-WHM1, S-PM2	(Wilson et al., 1993)
S-PWM1, S-PWM3	(Suttle and Chan, 1993)
S-BNM1, S-MM1, S-MM4, S-MM5, S-RSM1, S-RSM2	(Wilson, 1994)
S-BP3, S-BM3	(Fuller et al., 1998)
S-IO9, S-IO10, S-IO14, S-IO17, S-IO18, S-IO21, S-IO25, S-IO26, S-IO36, S-IO41, S-IO42, S-IO43, S-IO46, S-IO47	(Clokie et al., 2006b)
S-IO48, S-IO50	pers. comm. A. Millard
Syn9	(Weigele et al., 2007)
Red Sea strains 1-89 except for strains 1, 2, 17, 21, 36, 47, 54, 63, 66, 71, 72, 78 and 86	(Millard and Mann, 2006)

3.2.5 Host range

The cyanobacterial strains listed in Table 3.2 were used to test the host range of phages A-HIS1 and A-HIS2. *Acaryochloris* strains were maintained as in § 2.2.2. *Synechococcus* strains provided by Nikki Wilkinson were maintained in ASW at 23 °C. Strain RS9917 was supplemented with 2 mM ammonium chloride.

Prochlorococcus strains were provided by Prof. David Scanlan and maintained in PCR-S11 (Rippka, 1988; Rippka et al., 2000; Moore et al., 2007). Well assays (§ 3.2.2) were performed and monitored for clearings for a week after the initial inoculation of phage and compared to a positive control.

Table 3.2: Cyanobacterial strains.

Genus	Strain	Reference
Acaryochloris	MBIC11017	(Miyashita et al., 1996)
	CCMEE5410	(Miller et al., 2005)
	HICR111A	pers. comm., R. Mohr
	A-HIS1R1, A-HIS1R2, A-HIS1R3, A-HIS2R1, A-HIS2R2	Phage resistant strains generated from strain MBIC11017 in this study (§ 5.2.1)
Synechococcus	CC9311	(Toledo and Palenik, 1997)
	WH7803, WH8102, WH8103, WH7805, WH5701, WH8109	(Waterbury et al., 1986)
	RS9909, RS9917	(Fuller et al., 2003)
Prochlorococcus	SS120 (CCMP-1375) MED4 (CCMP-1378)	(Chisholm et al., 1992)
	SB	(Shimada et al., 1995)
	GP2	(Shimada et al., 1995, 1996)
	MIT9312, MIT9313	(Moore et al., 1998)
	PCC9511	(Rippka et al., 2000)
	EQPAC1, NATL1MIT	(West et al., 2001)
Synechocystis	PCC6803	(Rippka et al., 1979)

3.2.6 Transmission electron microscopy

3.2.6.1 Phage sample preparation

Phages were concentrated from lysate by precipitation with 2 % (w/v) NaCl (Fisher Scientific Inc) and 10 % (w/v) polyethylene glycol, PEG-6000 (BDH) overnight at 4 °C in the dark followed by centrifugation at 11,000 g for 10 min. Phage pellets were resuspended in 2–3 mL 1× ASW and an equal volume of chloroform was added to remove the PEG. The sample was then shaken and centrifuged in a Hettich Rotina 46R centrifuge at 4,754 g for 15 min at 4 °C. The aqueous layer was removed from the chloroform layer and added to aqueous caesium chloride (CsCl, Fisher Scientific Inc) to a final concentration of 0.75 g mL^{-1}.

The solutions were then transferred to Beckman Ultra-Clear™ centrifuge tubes (14×95 mm) and ultracentrifuged at ∼155,000 g in a SW40Ti rotor for 18 h at 4 °C in an Optima L-80 XP centrifuge. Bands were removed using a syringe and the concentrated phage were dialysed in dialysis tubing (size 3/MWCO 12–14,000 Da, company) for 1 h against 1 L 1× ASW twice to remove the CsCl.

3.2.6.2 Negative staining

Grids were handled with inverted Dumont tweezers (no.5). Agar Scientific 400 mesh Cu carbon film grids (50, Cat. no. S160-4) were glow-discharged with an Emitech K100X Glow Discharger (EM Technologies Ltd.) on a glass slide, shiny-side up. Phage samples were negatively stained with 1 % (w/v) uranyl acetate (UA), which was kept on ice and stored at 4 °C in the dark. Typically, the grid was held with the tweezers shiny-side up, 5 μL of phage sample was applied to the grid and left for 1 min. Whatman filter papers (90 mm diameter, Cat. no. 1001090) were used to remove the sample followed by staining with UA for 45 s and then the grid was dragged with the

tweezers on the filter paper to remove the excess UA. The grid was then imaged or stored in a dessicator in the dark.

3.2.6.3 Imaging

Phages were imaged using a JEOL 1200EX TEM. The desired grid was selected at low magnification ($< 10{,}000\times$) with the objective aperture out using the fluorescent screen. Before changing to higher magnifications, the objective aperture was inserted to reduce sample damage. Objects of interest were identified and focussed on the fluorescent screen before observation on the computer screen via DigitalMicrograph^{TM} (Gatan Inc.) and a 1K Gatan camera. Images were saved and then processed with ImageJ.

3.2.7 Plaque expansion determination

Plaque assays were performed on exponentially growing culture (§ 3.2.3; see § 2.2.3, host growth curves included in Appendix A, Fig. A.1). Lysates were previously titred to ensure that the number of plaques per plate would be minimal (around 1–10). Plates with too many plaques were discarded. Plaques were chosen on the basis of being a single plaque in a spacious region of the bacterial lawn far from other plaques.

Usually, the experiment was continued until a plaque within the set of plaques under consideration merged with another plaque. Plaques were measured after the plaques of both phages had appeared. Photographs of the plaque were processed in ImageJ to determine plaque diameters. Plates were sealed with Parafilm M^{TM} after 2 days. Cell density was estimated by considering the number of cells after concentration per plaque assay aliquot and the area of the Petri dish.

3.2.8 One-step growth

Before beginning the one-step growth experiment, phage lysates were titred using plaque assays. 1.25 L of MBIC11017 cells were grown and 150 mL of cells was transferred to a 500 mL conical flask per replicate and inoculated with phage to have a multiplicity of infection (MOI) of 0.1.

Immediately after the inoculation of phage, two 0.5 mL samples were removed from each replicate. One sample was centrifuged at 13,000 g for 5 min at 4 °C in the dark, after which the supernatant was transferred into another tube to be later used to assess free phage. The other 0.5 mL of sample was incubated with four drops of chloroform to be later used to assess total phage.

The cultures were then incubated at 28 °C for 5 min to allow phage to adsorb. Cultures were shaken at 70 rpm under 30 μmol photons $m^{-2}s^{-1}$. In order to synchronise phage infection, the infected cultures were centrifuged at 6,693 g for 15 min at 28 °C. 0.5 mL of each supernatant was kept to assess the number of unadsorbed free phage.

Following this, the pellets of infected cells were then resuspended in 150 mL of fresh 1$\times$ ASW warmed to 28 °C. Two 0.5 mL samples were then removed and processed as before every two hours from inoculation (time zero). All phage samples were stored at 4 °C in the dark. The number of viable phages were then counted by preparing dilution series for each sample collected during the time course and two appropriate dilutions were assessed at each time point by plaque assay.

3.2.9 Phage DNA extraction

Phages were concentrated by PEG-precipitation and treated with chloroform as for TEM (see § 3.2.6.1). Concentrated phage were DNase treated at a concentration of 0.25 U mL^{-1} with Ambion TURBO™ DNase for 1 h at 37 °C. DNA was extracted from phages: 1 mL of phage concentrate was

added to an equal volume of phenol (Fisher Scientific), vortexed, left to stand for 4 min and centrifuged at 10,000 g for 4 min. The top aqueous layer was removed to a new Eppendorf and an equal volume of phenol:chloroform (Fisher Scientific) (1:1) was added, followed by vortexing, leaving for 4 min and centrifugation for 4 min. The top aqueous layer was again removed to a fresh tube and an equal volume of chloroform:isoamyl-alcohol (Fisher-Biotech) (24:1) was added followed by vortexing, leaving for 4 min and 4 min centrifugation. Following this, the top layer was removed and 0.4 volumes of 7.5 M ammonium acetate was added. Chilled isopropanol was then added at a ratio of 1:1. The Eppendorf was then gently inverted to precipitate DNA and left on ice for 30 min and then centrifuged at 13000 g for 20 min at 4 °C. The supernatant was removed and 200 μL of chilled 70 % ethanol was added and left on ice for 10 min. The DNA pellet was then centrifuged at 13000g for 20 min at 4 °C. The ethanol was then removed and the pellet was dried in a dessicator. The dried DNA pellet was then resuspended in 50 μL of UPW. DNA concentration was assessed on a NanoDrop® ND-1000 spectrophotometer.

3.2.10 Restriction fragment length polymorphism

Restriction fragment length polymorphisms (RFLPs) were assessed using *Eco*RI, *Bam*HI and *Hind*III (Invitrogen). 1.4 μg of DNA was incubated in a final volume of 30 μL containing 3 μL restriction endonuclease and 3 μL of the corresponding 10$\times$ buffer (*Eco*RI and *Bam*HI buffer REact 3: 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$ and 100 mM NaCl. *Hind*III buffer REact 2; same as REact 3 except with 50 mM NaCl). Gel-loading buffer II (6$\times$: 0.25 % bromophenol blue, 0.25 % xylene cyanol and 15 % Ficoll (type 400) in water, (Maniatis et al., 1982a)) was added to the DNA samples and then resolved by gel electrophoresis on 0.75 % (w/v) agarose gels (ultra pure agarose (Helena) in 1$\times$ TBE (Tris/Borate/EDTA) buffer (89 mM Tris

borate, 89 mM boric acid and 2.0 mM Na_2EDTA), (Maniatis et al., 1982a)) containing 0.1 % (v/v) ethidium bromide which was ran for 2 h at 100 V in 1× TBE buffer (Biorad Power pac 300 power supply). DNA fragment sizes were quantified using the GeneRuler™ 1 kb DNA ladder (Fermentas). Gels were observed under UV and imaged in GeneFlash on a Syngene Bio Imaging transilluminator unit.

3.2.11 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) was used to assess phage genome size. 50 μL of purified phage lysate concentrate (PEG-precipitated and chloroform treated) was added to 50 μL of 2% (w/v) molten low melting temperature agarose (SeaPlaque® GTG® agarose, BioWhittaker Molecular Applications) in 1× TE (Tris/EDTA buffer; 10 mM Tris, 1 mM EDTA, pH 8.0). Agarose plugs were set in CHEF Mapper® disposable plug moulds (Bio-Rad Laboratories, Inc.) and then incubated in lysis buffer (100 mM EDTA, 100 mM Tris-HCl, 1 % SDS and 0.5 mg mL^{-1} proteinase K (fungal, Invitrogen), pH 9.0) overnight at 55 °C.

Each plug was then dialysed 3× against 1× TE for 1 h. The plugs were then loaded into a 1 % Pulsed Field Certified™ Agarose gel (Bio-Rad) together with two ladders (Sigma Pulse Marker™ 0.1–200 kb and Bioline 5 kb DNA MegaLadder). The gel was submerged in a 0.5× TBE buffer with 45 mM Tris borate and 1 mM EDTA cooled to 14 °C and ran at 6.0 V cm^{-1} for 24 h using a CHEF Mapper® XA System (Bio-Rad). The gel was then stained with ethidium bromide for 30 min. DNA bands were observed under UV and imaged in GeneFlash on a Syngene Bio Imaging transilluminator unit.

3.3 Results and discussion

3.3.1 Phage isolation

To optimise phage isolation techniques filtered sea-water samples from the Red Sea were tested for phages infecting strain MBIC11017. These samples all tested negative for phages. Having optimised the technique, sea-water samples from the reef flat North of Heron Island, Australia, a known habitat for *Acaryochloris* spp., were then tested (March 2006). Phage isolations were performed using non-axenic cultures of strains MBIC11017 and CCMEE5410. Cultures were made axenic using a combination of streak-plating and contamination plates (§ 2.2.8). Axenic cultures were used in subsequent experiments unless stated otherwise.

Plaque assays were successful and yielded two plaque-forming lytic bacteriophages on *A. marina*, strain MBIC11017 Fig. 3.2. Both bacteriophages came from the same crude unfiltered water sample obtained by incubating an ascidian (*Lissoclinum patella*) from the coral reef of Heron Island with SM buffer. Plaques assays performed with the other Heron Island samples were negative for phages for both strains MBIC11017 and CCMEE5410. The two lytic phages were made clonal and subsequently characterised.

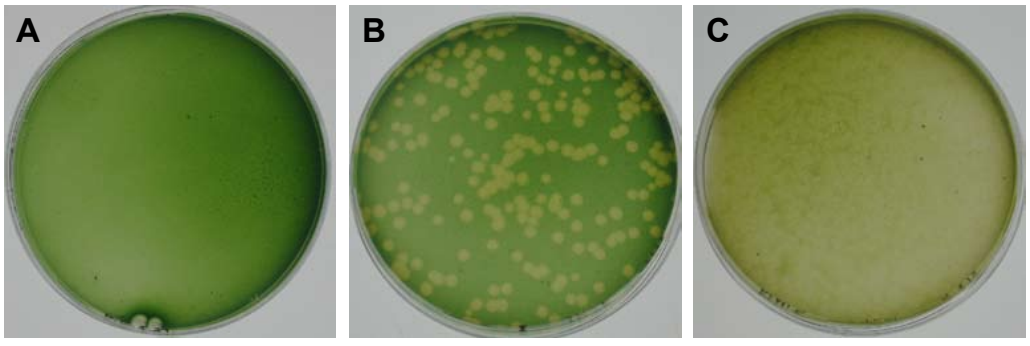


Fig. 3.2: *A. marina* plaque assays. (A) Negative control with ASW instead of seawater/virus sample, (B) positive plaque-forming units (p.f.u.s) or plaques and (C) confluent lysis.

A. marina strain MBIC11017 was also infected with the *Synechococcus* phages listed in § 3.2.4 in well assays which showed no clearings.

3.3.2 Phage morphology

TEM was used to examine phage samples. TEM was initially performed on crude samples from either plaque assays or lysate. However, grids were found to be covered with debris and phages were visible but difficult to image clearly. Subsequently phages were purified and concentrated using CsCl gradients, which was found to be ideal for image acquisition with phages immediately identifiable even at a low magnification (Fig. 3.3(D)). Depending on the phage preparation phage heads would often appear more round and collapsed (Fig. 3.3(C/F)) than icosahedral (Fig. 3.3(A)) and positively stained (Fig. 3.3(C/F)).

The two lytic phages isolated from plaque assays (§ 3.2.3) were examined and found to have icosahedral capsid heads and long, flexible, non-contractile tails. These features are characteristic of the phage family *Siphoviridae*. *Siphoviridae* are double-stranded DNA (dsDNA) viruses of the order *Caudovirales* or tailed phages.

The two phages were subsequently named A-HIS1 and A-HIS2 (A: *Acaryochloris*, HI: Heron Island, S: Siphovirus) and are shown in Fig. 3.3(A/B/C) and Fig. 3.3(D/E/F), respectively. For comparison, images of the myovirus S-PM2 were also obtained, to show the difference in size, morphology and structure as demonstrated by Fig. 3.3(K) and with its tail contracted and tail spike exposed in Fig. 3.3(L).

3.3.3 Tail structure

The main morphological difference between the two phages was tail length, with the tail of A-HIS1 shorter than that of A-HIS2 by $\sim$55 nm (Table 3.3). Both phages A-HIS1 and A-HIS2 had modular protein structures at the

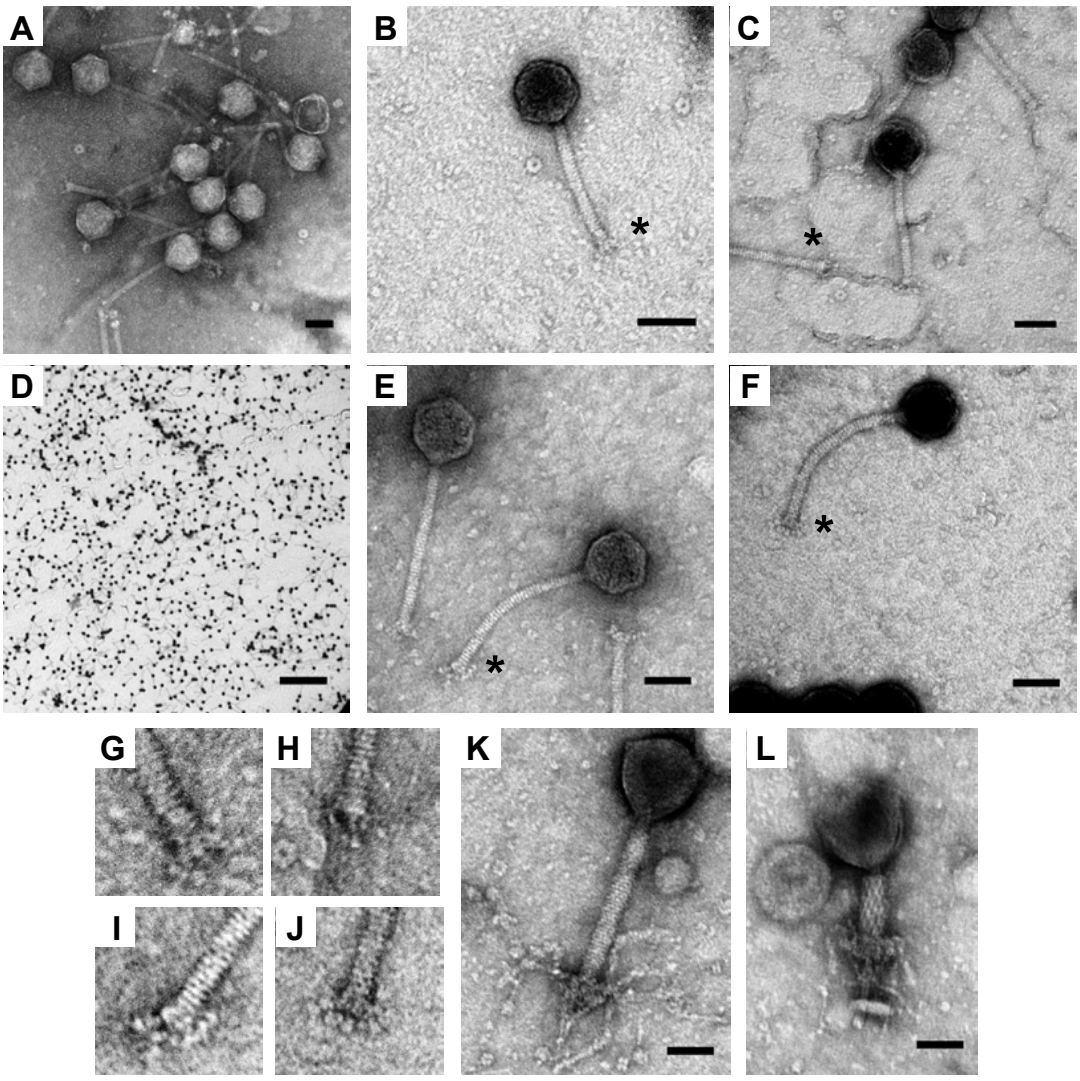


Fig. 3.3: Phage TEM. Bacteriophage morphology: electron micrographs of negatively stained *Acaryochloris* siphoviruses (A/B/C) A-HIS1 and (D/E/F) A-HIS2 on carbon film. (G/H/I/J) Detail of distal end of phage tail of phages in (B/C/E/F) marked with *, respectively. For comparison, (K/L) the *Synechococcus* myovirus S-PM2 with and without tail contraction. Note the exposed tail spike in (L). Scalebar: (A/B/C/E/F/K/L) 50 nm, (D) 1 μm. Magnification: (A) 120,000×, (B) 250,000×, (D) 10,000× and (C/E/F/K/L) 200,000×.

distal end of their tails similar to the receptor-binding protein structures characteristic of the temperate lactococcal bacteriophage TP901-1 (Spinelli et al., 2006). Like the capsid heads, the resolution of the receptor structure was dependent on the particular phage preparation and were often completely absent or partially degraded Fig. 3.3(H).

Table 3.3: Phage measurements. Measurements are the averages ($\pm$ one standard deviation) obtained from 10 individual phages for A-HIS1 and A-HIS2, imaged at 200,000$\times$ magnification. Tail lengths were measured from where the head joins the tail to the end of the receptor structure.

Phage	A-HIS1	A-HIS2
Measurement (nm)		
Head height	62.8$\pm$2.8	65.6$\pm$3.0
Head width	61.7$\pm$1.7	61.3$\pm$2.2
Tail length	124.1$\pm$3.8	178.9$\pm$6.1
Tail width	9.8$\pm$0.5	10.1$\pm$0.5
Tail subunit height	1.5$\pm$0.1	1.6$\pm$0.2

3.3.4 Phage classification

In 1998, Maniloff and Ackermann published a taxonomy classification on the *Caudovirales*, which divided the *Siphoviridae* into six genera: λ, T1, T5, L5, c2 and ψM1-like (Maniloff and Ackermann, 1998). A-HIS1 and A-HIS2 do not appear to fall into any category based on the receptor morphology from the TEM Fig. 3.3(G/H/I/J). In 2001 Brüssow and Desiere proposed two new *Siphoviridae* genera Sfi21-like and Sfi11-like and listed four other possible genera: sk1, ϕC31, TM4 and SPβ-like (Brüssow and Desiere, 2001).
To date, the *Siphoviridae* as classified by the International Committee on Taxonomy of Viruses (ICTV) have been expanded only by the addition of the new genera: N15-like and ϕC31-like, leaving a vast number of siphoviruses unclassified as of yet including lactococcal phage TP901-1.
Indeed, TEM has been established as the primary method to identify and classify phages, although it is not always possible to clearly identify the finer phage protein structures such as small tail fibres even with the best

of preparations. In light of these observations it would appear premature to place A-HIS1 and A-HIS2 into any one particular *Siphoviridae* genus at this time. The use of comparative genomics in phage taxonomy is discussed in § 4.3.6.

3.3.5 Plaque expansion

Plaque assays of phages A-HIS1 and A-HIS2 indicated that their plaques were able to grow indefinitely at the conditions they were kept at, only limited by the boundaries of the Petri dish. It was observed from plaque assays that after inoculation at the same time, A-HIS1 plaques appeared after A-HIS2 plaques. Both phages displayed clear circular plaques and the plates were sealed with Parafilm M^{TM} to reduce evaporation.

To investigate this observation further plaque growth was measured, which showed that the plaque radii for both phages grew linearly with respect to time and despite the early appearance of A-HIS2 plaques, the rate of plaque growth was similar for phages A-HIS1 and A-HIS2 Fig. 3.4. Continuous phage plaque growth in bacterial lawns has been observed in other phages, namely the podoviral coliphage T7 (Yin, 1991, 1993).

The ability of A-HIS1 and A-HIS2 plaques to grow indefinitely in the biofilm-like conditions of the plaque assay, gives some idea of how these phages would behave in the natural environment, since *Acaryochloris* spp. have been found in communal biofilms (Kühl et al., 2005; Larkum and Kühl, 2005) or as epiphytic colonies on the thalli of red algae (Murakami et al., 2004). As such, this particular phage-host system provides a good basis for the study of phage-host dynamics in a marine biofilm, which would be ecologically relevant to the distribution and survival of *Acaryochloris* spp. in their natural habitats.

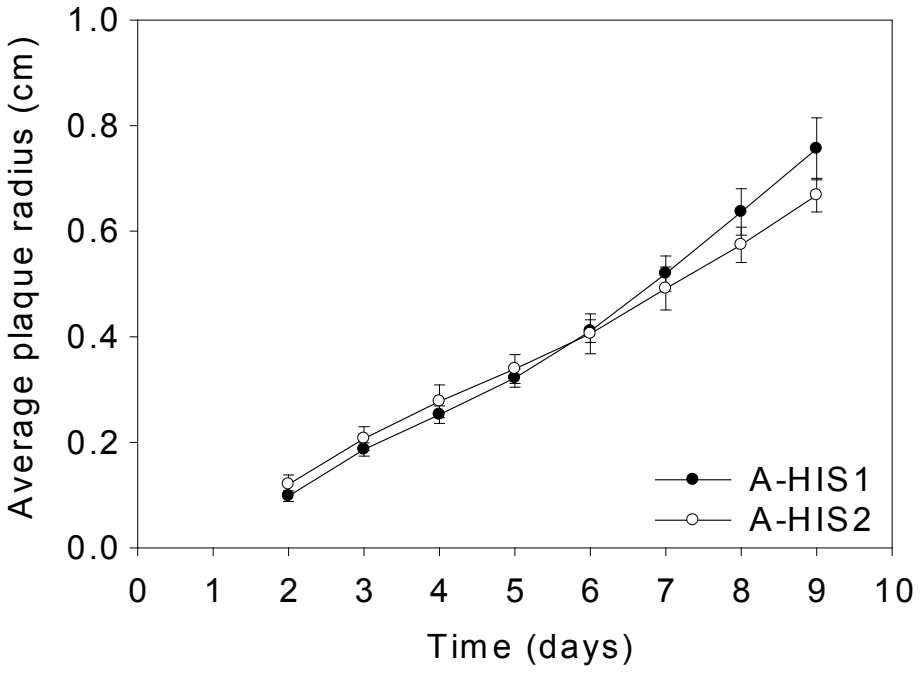


Fig. 3.4: Phage plaque growth. Cell density 0.175 cells $^{-1}\mu m^{-2}$. Error bars represent one standard deviation (n=12).

3.3.6 One-step growth curves

One-step growth curves were performed to determine phage growth parameters for the lytic phages A-HIS1 and A-HIS2. The parameters measured from this experiment characterise the lytic replication cycle of the phages, which allows the correlation of other observations over the infection cycle with the state of growth of the phages.

The average from two one-step growth curves with a 0.1 MOI (multiplicity of infection or phage:bacteria ratio) was analysed to determine growth parameters during the lytic infection cycle for phages A-HIS1 and A-HIS2 (Table 3.4, Fig. 3.5). The latent period, the time from infection to lysis was 5 h for both phages. The eclipse period (time taken for the first new viable phage to appear in the host after infection) for both phages was 3.25 h. The average burst size for A-HIS1 was 9.9 phages produced per infected bacterium compared to 19.1 for A-HIS2.

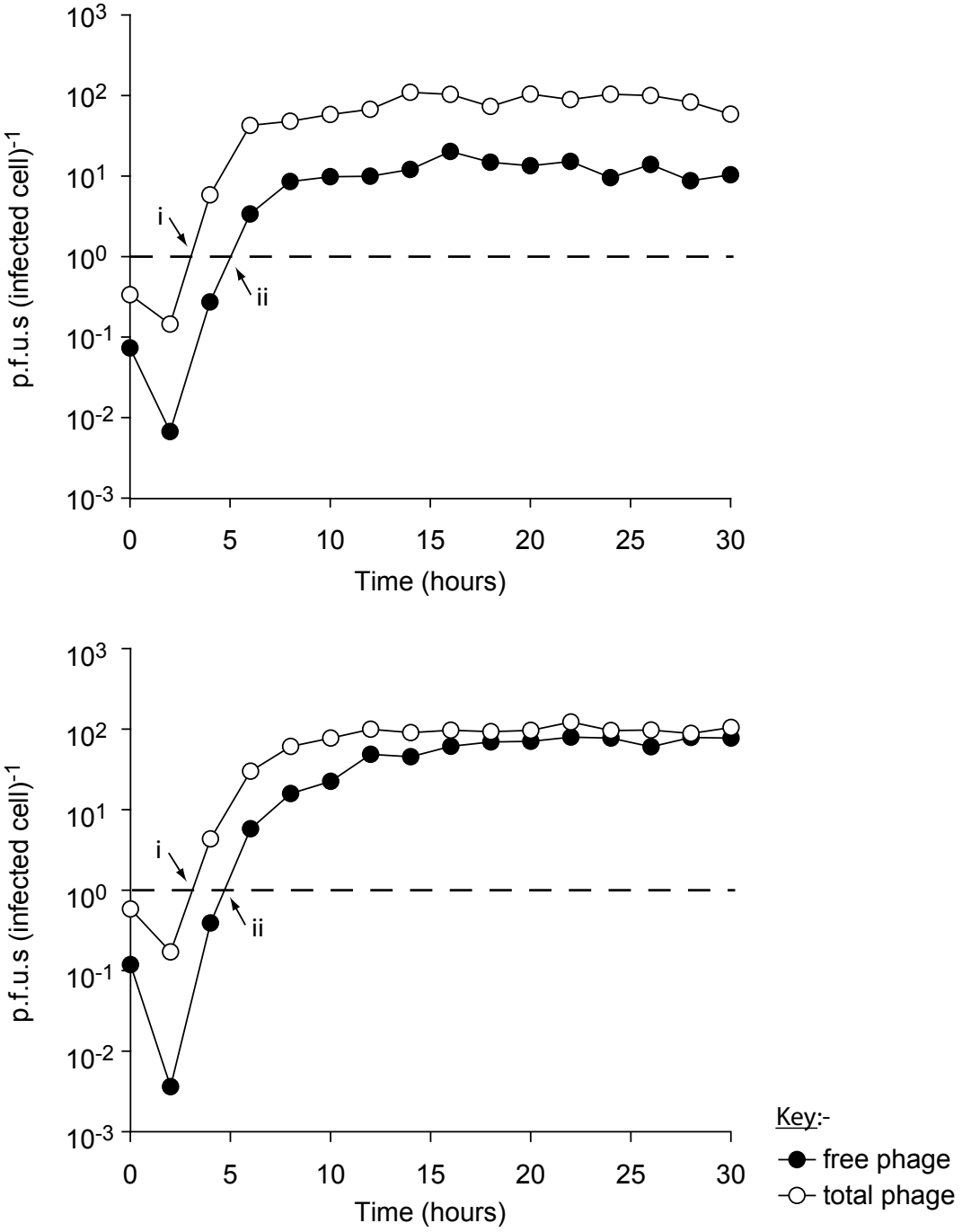


Fig. 3.5: Phage one-step growth experiments. One-step growth curves of (A) A-HIS1 and (B) A-HIS2 on strain MBIC11017. From time zero, the eclipse/latent period ends at i/ii where the total/free phage reaches one phage per infected cell, respectively. MOI = 0.1 and n=2 (error bars not included since a log scale is used). Free phage were removed between 0 and 2 h to synchronise infection (see § 3.2.8).

Table 3.4: Phage parameters. *The length of the eclipse and latent periods are the averages of two biological replicates, calculated from the inoculation time. Burst size was calculated as an average based on the 10 h and 12 h time points for A-HIS1 and the 8 h and 10 h time points for A-HIS2 of the two biological replicates. p.f.u.s = plaque forming units. Error of one standard deviation included in parentheses.

Phage	A-HIS1	A-HIS2
Length of eclipse period* (h)	3.25 (0.35)	3.25 (0.35)
Length of latent period* (h)	5 (0.00)	5 (0.71)
Burst size (p.f.u.s per cell)	9.9 (2.48)	19.1 (11.11)

Taking these results into account, it would seem that the reason A-HIS2 plaques appeared before A-HIS1 plaques (§ 3.3.5) was because A-HIS2 has a larger burst size since there is no difference between the eclipse or latent periods of the two phages. However, this does not explain why there is largely no difference between the plaque growth rates of A-HIS1 and A-HIS2 (Fig. 3.5). This is perhaps due to the limiting diffusion of the phages through the soft agar layer in a plaque assay despite the burst size of the phage.

Compared to other cyanophages which have had their growth properties characterised, A-HIS1 and A-HIS2 have relatively rapid replication cycles with the onset of lysis at 5 h. For instance, the *Synechoococcus* siphovirus S-BBS1 takes 9 h as does the myovirus S-PM2 (Suttle and Chan, 1993; Wilson et al., 1996). Even longer latent periods of 24–36 h have been shown for *Synechococcus* siphovirus isolates S-CBS2, S-CBS3 and S-CBS4 (Wang and Chen, 2008).

The burst sizes of the siphoviruses from these studies, which all infect *Synechococcus* spp., are greater than those reported here ($\sim$10–20) and vary from 57 for S-CBS4 to 250 for S-BBS1. The burst sizes and latent periods for phages A-HIS1 and A-HIS2 are consistent with the theory that shorter latent periods are selected for in higher densities of host bacterium (Abedon et al., 2001, 2003), since *A. marina* occurs within or underneath ascidians, locally in relatively higher concentrations than *Synechococcus* spp. which occupy the water column.

3.3.7 Host range

A-HIS1 and A-HIS2 were tested against 27 different cyanobacterial strains (§ 3.2.5) including the A-HIS1 and A-HIS2 phage resistant strains using well assays of which 26 produced no clearings in the wells. The two phages successfully infected and replicated *A. marina* strain MBIC11017, which they were isolated on. This suggests that phages A-HIS1 and A-HIS2 do not have a broad cyanobacterial host range. However, regarding *Acaryochloris* spp. it is worth noting that only three of the *Acaryochloris* genus were tested for susceptibility to phages A-HIS1 and A-HIS2. This suggests that the non-susceptible *Acaryochloris* strains HICR111A and CCMEE5410 may have diverged from strain MBIC11017 in terms of the components involved in phage-host recognition, e.g. adhesins. Plaque assays on strain CCMEE5410 were negative. It was not possible to perform plaque assays on strain HICR111A as the cells were inclined to clump in liquid culture and did not resuspend at all.

3.3.8 Restriction patterns and genome sizing

Like other members of the order *Caudovirales*, A-HIS1 and A-HIS2 both contain dsDNA as evidenced by the ability to generate restriction fragment length polymorphism (RFLP) patterns with enzymes which specifically cut dsDNA (Fig. 3.6). Underrepresented bands in the RFLP patterns may be due to incomplete enzyme digestion. On comparison of the RFLP patterns for any of the enzymes used, it was clear the RFLP patterns were specific to each phage and therefore the phages were distinct.

Genome sizes of phages A-HIS1 and A-HIS2 were determined by pulsed-field gel electrophoresis (PFGE) and found to be between 50 and 60 kb (Fig. 3.6). To date (May 2009), 221 *Siphoviridae* genomes have been sequenced and deposited in the NCBI database. The genome sizes of A-HIS1 and A-HIS2 are larger than the average and median genome size

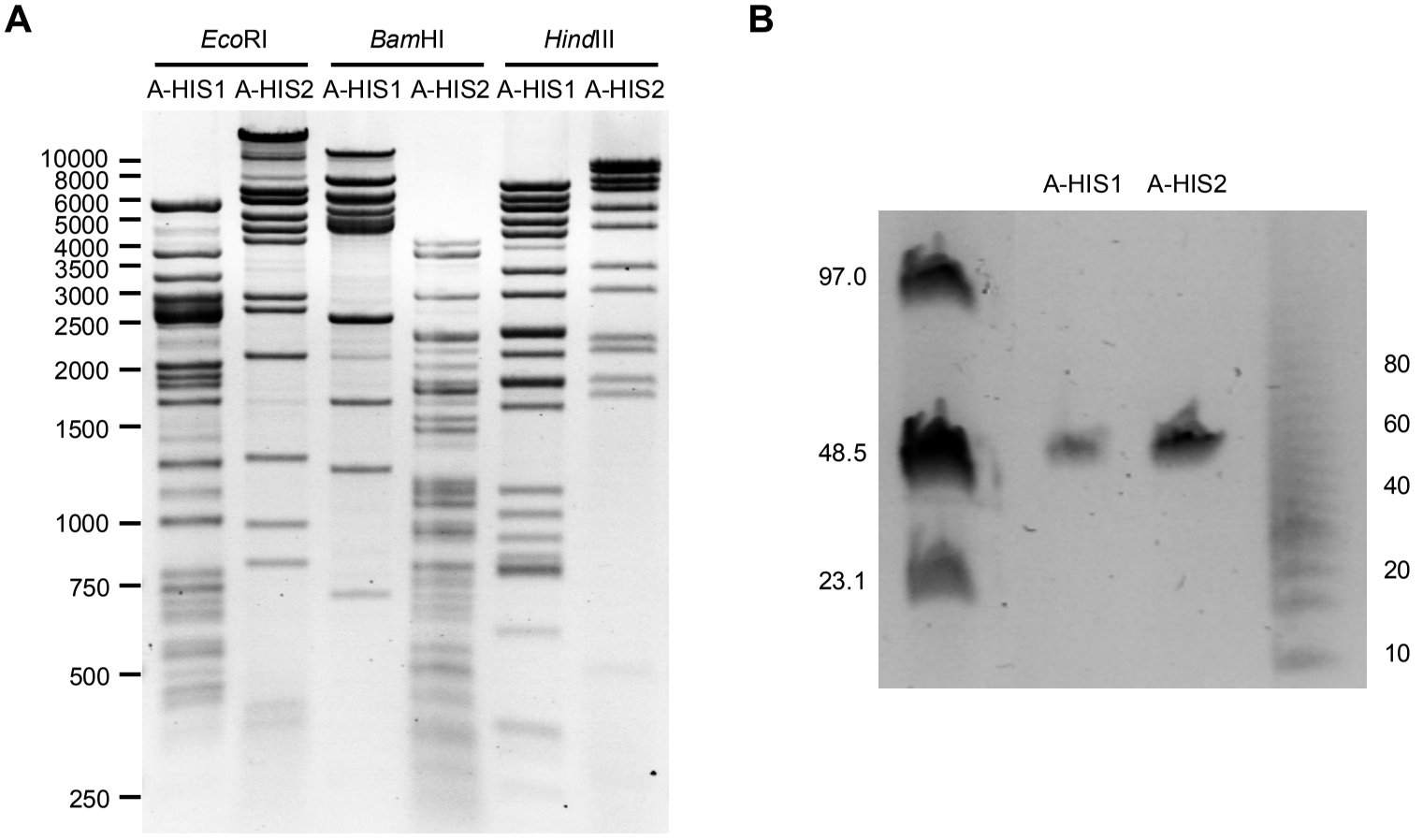


Fig. 3.6: RFLP and PFGE of purified phage genomic DNA. (A) Genomic DNA from phages A-HIS1 and A-HIS2 was digested with *Eco*RI, *Bam*HI and *Hind*III. (B) Phage genome size was assessed using two DNA ladders: Sigma Pulse Marker^{TM} 0.1–200 kb (left) and Bioline 5 kb DNA MegaLadder (right).

for *Siphoviridae* in the NCBI database which are 48.252 kb (to the nearest base) and 43.785 kb, respectively. The smallest and largest *Siphoviridae* genome sizes are 14.51 kb and 134.416 kb belonging to *Lactococcus* prophage bIL311 (Chopin et al., 2001) and *Bacillus* phage SPBc2 (Koonin, 2006), respectively.

3.4 Conclusions

A-HIS1 and A-HIS2 are the first phages of *A. marina* to be isolated and characterised. Notably, known marine cyanobacterial siphoviral isolates are rare in the literature; the few examples include *Synechococcus* spp. phages P1 (Lu et al., 2001; Zhong et al., 2002), S-BBS1 (Suttle and Chan, 1993), S-CBS2, S-CBS3, S-CBS4 (Wang and Chen, 2008), and *Prochlorococcus* sp. MIT9313 phages P-SS1 and P-SS2 (Sullivan et al., 2003). There are limited studies on the growth and behaviour of such cyanobacterial siphoviruses and therefore, this study gives an insight into the nature of these entities.

The plaque behaviour of phages A-HIS1 and A-HIS2 indicated that they are likely to proliferate in a biofilm-like environment. This virulent behaviour may confer phage resistance mechanisms to the host species, since strains resistant to phages A-HIS1 and A-HIS2 readily grew § 3.2.5, which are investigated and discussed in more detail in Chapter 5. Importantly, the life cycle growth parameters of phages A-HIS1 and A-HIS2 were established, which may serve as a basis for the interpretation and planning of further experiments investigating the interaction between phage and host in Chapter 5.

This phage-host system therefore provides a good basis to study phage-host dynamics in a biofilm, especially as it is ecologically relevant to the lifestyle of the *Acaryochloris* spp. and may have important implications concerning their survival in biofilms (with or without other organisms) in the natural

environment.

To add to the complexity, A-HIS1 and A-HIS2 presumably originated from an MBIC11017-like *Acaryochloris* sp(p). present in the Heron Island habitat since we are able to infect strain MBIC11017, which originates from the coast of Palau (Miyashita et al., 1996). Therefore, it is not unreasonable to assume that such phages may be found in the *Acaryochloris* habitat in Palau, and that the *Acaryochloris* spp. of both Heron Island and the coast of Palau are considerably similar, at least in terms of phage receptors, despite the distance they are apart.

Significantly, strain CCMEE5410 is not affected in the presence of these phages in both well and plaque assays (investigated further in Chapter 5). In Chapter 2, it was shown that strains MBIC11017 and CCMEE5410 are significantly different from each other due to the absence of accessory light harvesting antennae in strain CCMEE5410 (Miller et al., 2005), which may have been seeded as contamination from fishing stocks into the Salton Sea (Ritchie, 2006). Comparison of the genomes of strains MBIC11017 (Swingley et al., 2008) and CCMEE5410 may provide more insight into the divergence of these strains due to adaptive radiation in their ecological niches.

Chapter 4

Genomic and proteomic characterisation of *Acaryochloris* phages A-HIS1 and A-HIS2

4.1 Introduction

Viruses are a large source of genetic diversity and have a great effect on the evolution of species through mechanisms such as LGT (§ 1.3.2), and the influence they may have on the physiology of their hosts (Clokie and Mann, 2006). Their enigmatic existence has posed a great problem for taxonomists and evolutionary biologists since no unified method has been able to fully account for all the different aspects of their genetic and morphological signatures. This chapter details a genomic and proteomic characterisation of phages A-HIS1 and A-HIS2. These are the first genomes to be sequenced of phages which infect *A. marina* and to date of any siphoviruses. The presence of novel genes in these phages sheds some light into the origin of mitochondrial DNA polymerase (mtDNA pol) γ, which has long puzzled evolutionary biologists.

4.2 Materials and methods

4.2.1 Genome sequencing and annotation

DNA was extracted from CsCl-purified phages and stored in 70 % (v/v) ethanol on ice (§ 3.2.6.1, § 3.2.9). Phage genomes were commercially sequenced by AGOWA (LGC) using a shotgun approach. Phage DNA was sonicated to 2–3 kB fragments and the ends were polished with T4 DNA polymerase/T4 polynucleotide kinase and fragments were ligated into pMCL200. Clones were end-sequenced from both sides to get a 6 $\times$ coverage using standard Dye Terminator Sequencing (BigDye version 3.1) on a 96 capillary 3730XL DNA Analyzer (Applied Biosystems). Remaining gaps were closed by walking reads on shotgun clones or on PCR products spanning gaps. Assembly was done in GAP4 (Staden Package). No gap was left, that was not covered either by a shot gun clone or by a PCR product for both phage genomes, therefore the genome DNA molecules are circular. The genomes were annotated using Artemis (Rutherford et al., 2000). Genome maps were drawn with custom software written in the Q script language version 6 by Peter Wheatley (and edited by the author) similar to that of S-PM2 in (Mann et al., 2005). Specifically, the raw G+C data were smoothed by convolving with a Gaussian mask (σ=100 bases). Region annotation was performed separately.

4.2.2 Gene predictions and analyses

ORFs (open reading frames) were predicted using both GeneMark.hmm 2.0 (Besemer and Borodovsky, 1999) and GLIMMER 3.02 (NCBI) (Salzberg et al., 1998). To create the final set of predicted ORFs for each genome, the two sets of predicted ORFs from GeneMark and GLIMMER were first combined. For ORFs predicted by both programs where there was almost complete overlap, the longer of the two was kept. NCBI blastp and psi-blast

were used to assign putative function to the predicted ORFs (Altschul et al., 1990). ORFs were designated virion structural proteins (VSPs) from mass spectrometry analyses of purified phage proteins (§ 4.2.7). The comparative genomic analyses were used to further assign function to ORFs by virtue of the similarity between ORFs of the two phages (§ 4.2.3).

4.2.3 Genome comparisons

Comparison files were generated using standalone blast (tblastx or blastp) and genomes were visually compared using ACT (Carver et al., 2005). Gene similarity was assessed between the phage genomes by standalone blastp (default parameters) of each set of phage genes (amino acid sequences) first as query against the other, then as the subject (i.e. database) for the other. Results with e-value $> 1e^{-3}$ were removed. Shared gene pairs were the top hits in both blastp searches which could mutually identify one another. Similar gene pairs were those which could also mutually identify one another, but were not top hits in the blastp searches. The % identity value between two ORFs was obtained from a percent identity matrix (PIM) calculated from the two ORF amino acid sequences in ClustalX (1.83) using a complete alignment with default parameters (gap opening, gap extension): pairwise (10, 0.1) and multiple (10, 0.2) (Thompson et al., 2002). Typically, for comparison, gap opening/extensions penalties for both pairwise and multiple alignment were set to 2 and 0.1, respectively. The genomes were also compared visually using Dotter (Sonnhammer and Durbin, 1995).

4.2.4 Motifs and regulatory elements

The programs used to detect motifs and regulatory elements are as listed in Table 4.1. A custom Perl script for handling ELPH output files was co-written by Antony Holmes and Andrew Millard. For ribosomal binding sites (RBS), the region length was set to 50 specified with the sequences

AGGAGG and GGAGAA, respectively derived from the N-terminal 16S rRNA sequences rrn16Sa (NCBI GeneID: 5680242) and rrn16Sb (NCBI GeneID: 5684370) from the *A. marina* genome Swingley et al. (2008). Spacers of CRISPRs (clustered regularly interspaced short palindromic repeats) were obtained from the CRISPR database (Grissa et al., 2007b). The spacers were blasted against the phage genomes using standalone blastn with gap penalty settings: existence = 1 and extension = 2. Slippery sequences were detected using Programmed Frameshift Finder[1] with default settings.

Table 4.1: Motif and regulatory element prediction programs.

Genome feature	Program	Reference
tRNAs	tRNAscan-SE 1.21	(Lowe and Eddy, 1997)
transcriptional terminators	transtermHP	(Kingsford et al., 2007)
RBS	RBSfinder	(Suzek et al., 2001)
CRISPRs	CRISPRFinder	(Grissa et al., 2007a)
promoters/motifs	PHIRE	(Lavigne et al., 2004)
	ELPH (Gibbs sampler)	(Pertea et al., 2007)
programmed frameshifts	Programmed Frameshift Finder	(Xu et al., 2004)

4.2.5 Metagenomic database searches

Amino acid sequence of the predicted phage ORFs were submitted to the CAMERA[2] database (Seshadri et al., 2007) and compared to the sequences in the metagenomics ORF peptides database using blastp. The peptide sequences with metadata were then downloaded via a spreadsheet and the corresponding read data to each sequence was collated, and then similarly the corresponding assembly data to each read was collated. The assembly duplicates were then removed and metadata was appended to the resulting list of assemblies. Following this, the assemblies were processed using Gen-

[1] http://chainmail.bio.pitt.edu/~junxu/webshift.html

[2] http://camera.calit2.net/, Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis

eMark.hmm 2.0 (Besemer and Borodovsky, 1999) and ORF Finder (NCBI) (and GLIMMER 3.02 (NCBI) (Salzberg et al., 1998) if there were any discrepancies between the ORFs detected by the other methods and the predicted ORFs provided in CAMERA). Subsequently, NCBI blastp was used to identify the assembly ORFs. The IP of Panama, Ecuador, Fr. Polynesia, the USA, Italy, Tanzania, Bermuda and Canada as the countries of origin of the read, assembly and protein data obtained from CAMERA, which were used in this study. Sequences from international waters (1) between Madagascar and South Africa and (2) 500 miles west of the Seychelles in the Indian Ocean from CAMERA were also used.

4.2.6 Phylogenetics

Alignments were created in Clustal X 1.83 (Thompson et al., 2002). Depending on the gene of interest, sequences for alignment were selected by combining the phage sequences with top blast hits, related sequences of interest (including potential outgroups for creating phylogenetic trees) and/or environmental hits from the CAMERA database (Seshadri et al., 2007). Typically, similar amino acid sequences were aligned (e.g. DNA polymerase I-type and DNA polymerase γ-type sequences were aligned separately), followed by profile alignment of the separate alignments. To improve the quality of the alignment, gap opening and extension parameters were respectively 2 and 0.1 in both the pairwise and multiple sequence alignment parameters to encourage gaps to form. Default alignment parameters of Clustal X were used to align gap poor sequences. The alignments were then truncated at each end to eliminate gap rich regions. Alignment adjustment by eye was performed in BioEdit when necessary. Alignments were then saved in nexus format in Clustal X and exported from Mesquite 2.01 for MrBayes with `interleave=yes`. Phylogenies were computed by MrBayes 3.1.2 using a Poisson model and a gamma-distributed rate varia-

tion across sites with a proportion of invariable sites (`rates=invgamma`), 500,000 generations, sampling frequency of 100 and a burnin of 1250 (25 %) (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Trees were visualised in Treeview 1.6.6 (Page, 1996). In the case of the DNA polymerase A domain analysis, the alignment was based on that proposed[3] by Filée et al. (Filee et al., 2002). The alignment of each sequence was checked comparatively by observing the NCBI conserved domain alignment of that sequence.

4.2.7 Proteomics

Phages were purified as for TEM (§ 3.2.6.1). Following dialysis, phage ghosts were prepared as described previously (Konopa and Taylor, 1979; Clokie et al., 2008). Protein samples were prepared for SDS-PAGE as described previously (§ 2.2.15, (Chan et al., 2007)). Protein samples were then resolved on a 15 % SDS-PAGE gel as described in § 2.2.15. Proteins were sequenced by MALDI-TOF (matrix-assisted laser desorption/ionisation-time of flight) and LC-MS/MS. Protein bands were identified by comparing the sequence data to predicted protein sequences acquired from the genome data. Molecular weight was assessed on the SDS-PAGE using the Fermentas PageRuler™ Prestained Protein Ladder Plus. PSORT-B was used to predict protein localisation (Gardy et al., 2003).

4.3 Results and discussion

4.3.1 Genome properties

Genome sizes determined by genome sequencing were 55,653 bp and 57,391 bp for A-HIS1 and A-HIS2, respectively, which corresponded to the estimates obtained by PFGE (§ 3.3.8). The circular[4] genomes of phages A-HIS1

[3] `http://www-archbac.u-psud.fr/Projects/dnapol/Ali_polA.htm`
[4] see § 4.2.1

and A-HIS2 are the first *Siphoviridae* genomes sequenced from phages which infect a cyanobacterium (Fig. 4.1, Fig. 4.2). Both phages have a similar average molecular G+C content of 47.1 and 47.2 %, respectively, which is close to the average molecular G+C content of the host strain of *A. marina* MBIC11017 at 47.0 % (Swingley et al., 2008).

4.3.2 Gene predictions

GeneMark predicted 83 ORFs for each phage and GLIMMER predicted 93 and 104 ORFs for A-HIS1 and A-HIS2, respectively. By analysing the two sets of predicted ORFs, 95 and 104 putative ORFs were assigned to A-HIS1 and A-HIS2, respectively (see § 4.2.2 for details). All ORFs were subjected to a blastp analysis (May 2009) against the nr database, where hits were deemed insignificant for e-value $> 1e^{-3}$ (Table 4.2 and Table 4.3). This analysis allowed 16 % of A-HIS1 ORFs and 13 % of A-HIS2 ORFs to be assigned a putative function (ORFs assigned as hypothetical proteins are included in Table A.1 and Table A.2, Appendix A). A-HIS1 ORF 38 and A-HIS2 ORF 59 had significant blastp hits. However, the hits were mostly hypothetical and so no putative function was assigned to these ORFs. Iterative psi-blasts did not yield any new information.

4.3.3 Transfer RNAs and pseudogenes

The *Acaryochloris* phage genomes have no pseudogenes and no transfer RNA (tRNA) genes. This is not unusual. Enterobacteria phages HK97 and T5 are the only siphoviruses which have pseudogenes out of the $\sim$221 genomes which have been sequenced (May 2009); both containing a single pseudogene. $\sim$10 % of the sequenced siphovirus genomes contain between 1 and 5 tRNA genes, with the exception of T5 and Mycobacterium phage Wildcat, which have 25 and 22 tRNA genes, respectively.

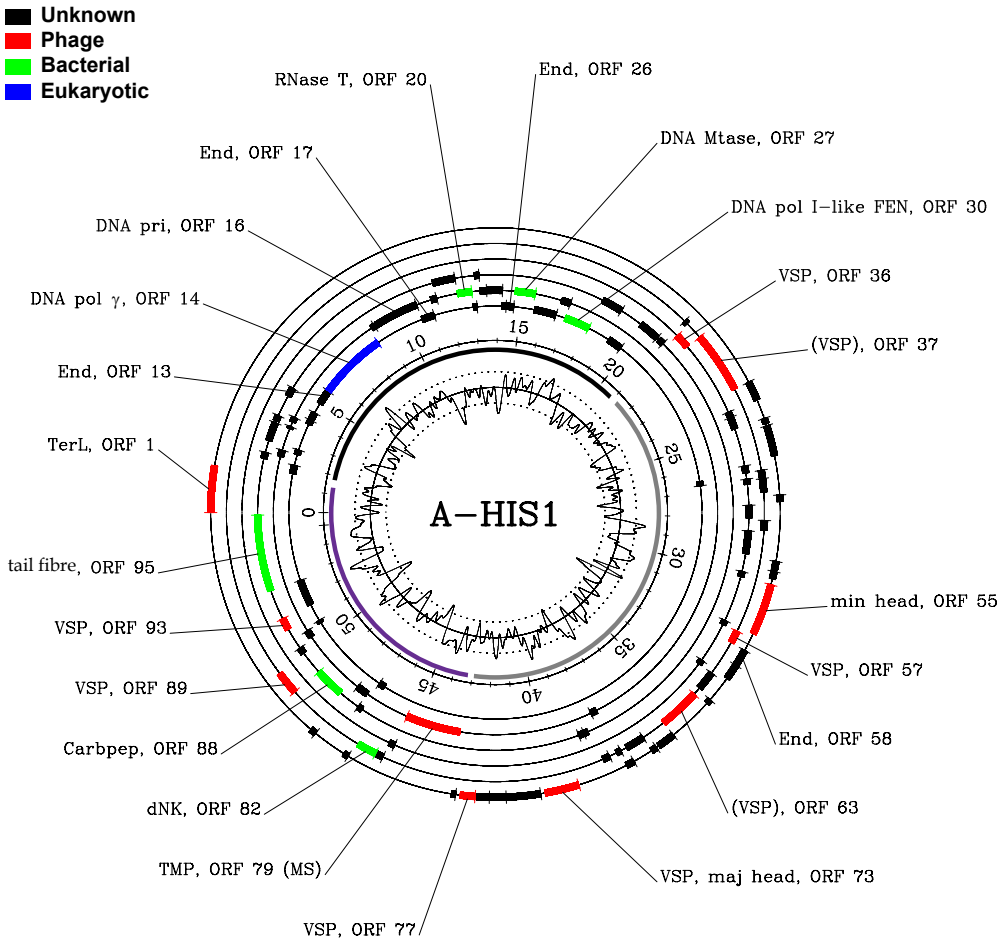


Fig. 4.1: Organisation of the A-HIS1 genome. Starting from the centre of the genome map is the molecular G+C content (solid line 40%, dotted lines 35 % (inner) and 45% (outer)). Next are the regions as defined by the author; region I (black), II (grey) and III (purple). This is followed by the scale (in kilobases) and the final six circles represent the six reading frames. All ORFs shown code for predicted proteins. The ORF end with protrusions indicates the 5′ end of the ORF. VSP indicates a virion structural protein which was identified by mass spectrometry. (VSP) denotes an ORF categorised as a putative A-HIS1 phage structural protein based on amino acid sequence similarity to a corresponding ORF in A-HIS2 identified as a putative VSP by mass spectrometry and vice versa. (MS) denotes an ORF which was identified by both blastp and mass spectrometry. See Table 4.2 and Table 4.3 for gene details. Original figure generated by Peter Wheatley and edited by the author.

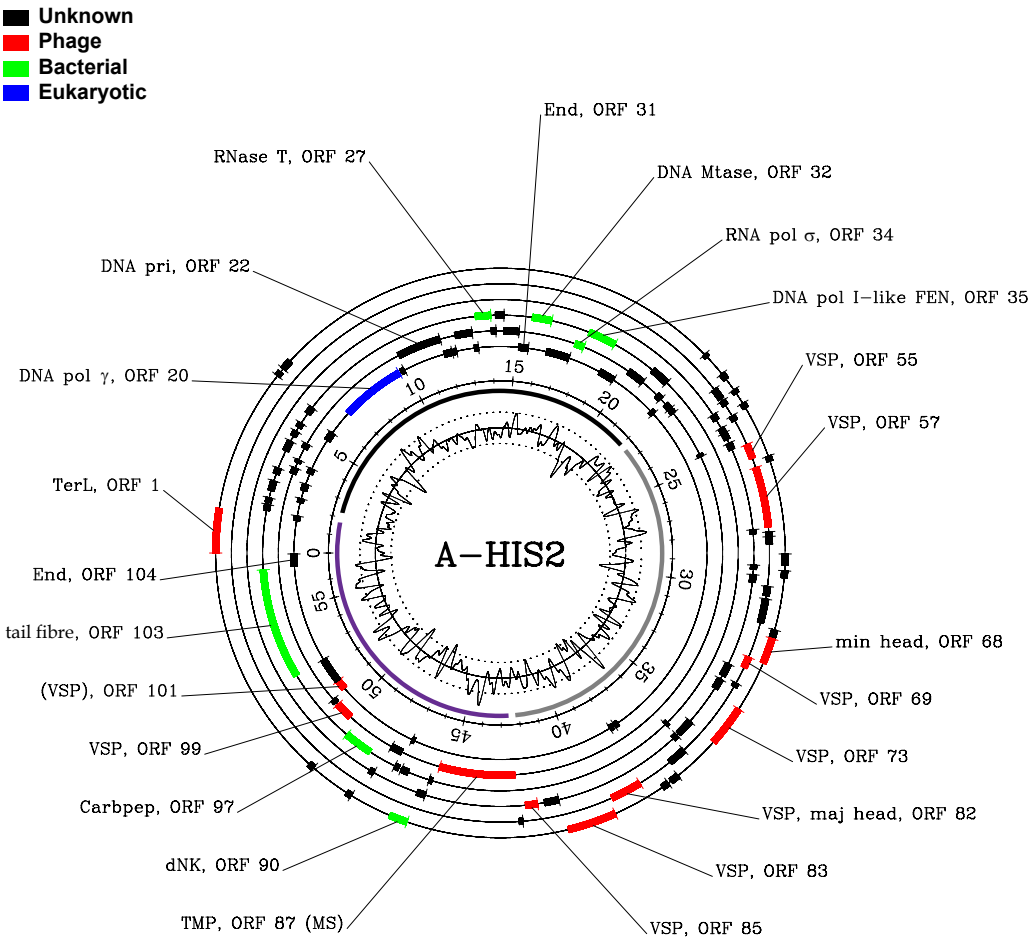


Fig. 4.2: Organisation of the A-HIS2 genome. Details as for Fig. 4.1. Original figure generated by Peter Wheatley and edited by the author.

Table 4.2: Putative function of A-HIS1 ORFs. Predicted ORFs from the genome of A-HIS1 with identified putative function identified by mass spectrometry* (MS) or blastp.

ORF	Start	End	Strand	Size (aa)	E-value	Conserved domains	Comment
1	1	1461	+	486	2e-17	Terminase 6, terminase-like family, multi-domain, pfam03237 (2e-13)	TerL, terminase large subunit. Top 2 hits bacterial which contain domain, top phage hit *Streptococcus* phage EJ-1.
13	4916	5563	-	215	7e-08	AP2, DNA-binding domain in plant proteins such as APETALA2 and EREBPs, non-specific hit, smart00380 (3e-04) (AP2 superfamily)	End, endonuclease, top hit pathogenesis-related transcriptional factor and ERF protein (*Burkholderia multivorans* ATCC 17616. Also phage and eukaryota hits.
14	5573	8554	-	993	1e-107	DNA polymerase family A, pfam00476 (4e-25), non-specific hit, DNA_pol_A superfamily	DNA pol γ, DNA polymerase γ similar to eukaryotic mitochondrial DNA polymerase γ. Top hit fungal *Botryotinia fuckeliana* B05.10
16	8883	10700	-	605	-	Phage associated DNA primase, multi-domain, COG3378 (2e-03)	DNA pri, putative DNA primase
17	10700	11248	-	182	2e-16	AP2 domain, pfam00847 (1e-03), non-specific hit, AP2 superfamily	End, top hit numod4 motif/putative phage protein *Clostridium botulinum* Ba4 str. 657
20	12390	12965	-	191	3e-12	DEDDh exonucleases, specific hit, DnaQ_like_exo superfamily cd06127 (6e-17)	RNase T, putative bacterial ribonuclease T homologue, top hit gamma proteobacterium NOR51-B, also hits to DNA polymerase III
26	14372	14698	-	108	8e-04	AP2 domain, non-specific hit, pfam00847 (2e-03)	End, similar to p42.1 *Xanthomonas* phage Xop411
27	14703	15512	-	269	2e-04	AdoMet_MTases, S-adenosylmethionine-dependent methyltransferases, cd02440 (2e-04)	putative DNA modification methylase similar to *Pseudomonas* sp. ND6 DNA methylase
30	16964	18061	-	365	2e-08	53EXOc, T5 type 5-3 exonuclease, cd00008, (9e-13), XPG superfamily	DNA pol I-like flap endonuclease (FEN), top hits *Finegoldia magna* ATCC 29328 DNA polymerase I, mostly bacterial DNA polymerase I hits
36*	20997	21506	+	169	-	-	VSP, virion structural protein

37	21542	23608	+	688	-	-	VSP based on similarity to AHIS-2 ORF 57 identified by MS; see Table 4.5
38	23614	24228	+	204	4e-52	-	Phage-related protein, top hit *Acinetobacter baumanii* AYE, similar to *Burkholderia* phage Bcep1 gp12 and cyanophage P-SSP7 protein PSSP7_001, also hit to Dcm coliphage TLS
55	30097	31716	+	539	1e-09	Phage Mu protein F like protein, non-specfic hit, pfam04233 (3e-10), Phage_Mu_F superfamily	min head, minor head protein , top hit phage phiJL001
57*	31871	32284	+	137	-	-	VSP
58	32302	32832	+	176	2e-12	AP2 domain, pfam00847 (3e-07), non-specific hit, specific hit cd00018 (3e-06), AP2 superfamily	End, top hit p42.1 *Xanthomonas* phage Xop411, other hits include HNH endonucleases, bacterial, phage and eukaryota
63	34385	35767	+	460	-	-	VSP based on similarity to AHIS-2 ORF 73 identified by MS; see Table 4.5
73*	39109	40188	+	359	-	-	VSP, maj head, major capsid protein
77*	42373	42843	+	156	-	-	VSP
79*	43158	45371	-	737	7e-20	Phage-related minor tail protein, pfam10145 (2e-13), non-specific hit, PhageMin_Tail superfamily	TMP, phage tail tape measure protein, top hit Nostoc punctiforme PCC 73102
82	45830	46480	+	216	2e-12	dNK, deoxyribonucleoside kinase, cd01673 (8e-19), non-specific hit, P-loop NTPase superfamily	dNK, deoxynucleoside/deoxyguanosine kinase, top hit Flavobacteria bacterium MS024-3C
88	48034	49149	-	371	6e-23	Peptidase subfamily C39A domain, cd02549 (1e-06), non-specific hit, Peptidase_C39_like superfamily	Carbpep, carboxypeptidase top hit *Lyngba* sp. PCC8106. Cyanobacterial and proteobacterial hits
89*	49193	49975	+	260	-	-	VSP
93*	51109	51552	-	147	-	-	VSP
95	52690	55557	-	955	8e-06	-	putative tail fibre protein, top hits are *Verrucomicrobia* proteins (immunoglobulin I-set domain and fibronectin type III)

Table 4.3: Putative function of A-HIS2 ORFs. Predicted ORFs from the genome of A-HIS1 with identified putative function identified by mass spectrometry* (MS) or blastp.

ORF	Start	End	Strand	Size (aa)	E-value	Conserved domains (e-value)	Comment
1	1	1437	+	478	1e-17	Terminase 6, terminase-like family, (multi-domains) pfam03237 (1e-12)	TerL, top hit *Elusimicrobium minutum* PBSX family phage terminase large subunit
20	6792	9671	-	959	3e-127	DNA polymerase family A, pfam00476 (4e-31), non-specific hit, DNA_pol_A superfamily	DNA pol γ, top hit *Paracoccidioides brasiliensis* Pb18
22	9997	11829	-	610	-	Phage associated DNA primase, multi-domains COG3378 (5e-04)	DNA pri, putative DNA primase
27	13373	13972	-	199	2e-10	DEDDh exonucleases, cd06127 (4e-12), specific hit, DnaQ_like_exo superfamily	RNase T, putative bacterial ribonuclease T homologue, top hit *Marinobacter algicola*, proteobacterial hits
31	15153	15578	-	141	1e-04	AP2 domain, pfam00847 (2e-05) non-specific hit, AP2 superfamily	Similar to p42.1 *Xanthomonas* phage Xop411
32	15578	16333	-	251	3e-04	Putative methyltransferase PHA02056 (6e-04), non-specific hit, NADB_Rossmann superfamily	DNA Mtase, similar to DNA N6-adenine methyltransferase, coliphage rv5
34	17476	17859	-	127	8e-04	-	RNA pol σ, RNA polymerase σ factor similar to *Ruegeria* sp. TM1040 ECF subfamily RNA polymerase σ-24 subunit
35	17843	18928	-	361	1e-08	53EXOc, 5-3 exonuclease, cd00008, (6e-14), non-specfic hit, XPG superfamily	DNA pol I-like FEN. Similar to *Finegoldia magna* ATCC 53516 DNA-directed DNA polymerase
55*	24908	25414	+	168	-	-	VSP
57*	25748	27811	+	687	-	-	VSP
59	27980	28135	+	51	1e-05	-	Similar to cyanobacterial protein, in particular *Acaryochloris marina* MBIC11017 AM1_2888. Top hit *Cyanothece* sp. PCC 7425
68	31483	32319	+	278	6e-09	Phage Mu protein F like protein, non-specfic hit, pfam04233 (7e-08)	min head, similar to SPP1 gp7 (*Ruegeria* sp. TM1040) and gp62 phage phiJL001
69*	32321	32707	+	128	-	-	VSP
73*	33994	35325	+	443	-	-	VSP
82*	38069	39145	+	358	-	-	VSP, maj head, major capsid protein
83*	39253	40845	+	530	-	-	VSP

85*	41700	42140	+	146	-	-	VSP
87*	42451	45576	-	1041	1e-21	PhageMin_Tail, phage-related minor tail protein, pfam10145 (8e-10), PhageMin_Tail superfamily	TMP, top hit *Nostoc punctiforme* PCC 73102, top phage hit *Clostridium* phage phi3626
90	46099	46698	+	199	2e-13	dNK, deoxyribonucleoside kinase, cd01673 (2e-18), non-specific hit, P-loop NTPase superfamily	dNK, top hits *Deinococcus radiodurans* R1 deoxyguanosine/adenosine kinase subunit and Bacteroidetes *Dyadobacter fermentans* DSM 18053
97	48341	49456	-	371	8e-25	C1 peptidase (papain) family, cd02619 (1e-06), non-specific hit, Peptidase_C1 superfamily	Carbpep, to hits alphaproteobacterial *Ochrobactum anthropi* ATCC 49188 and Lyngba sp. PCC8106 carboxypeptidase
99*	49825	50544	-	239	-	-	VSP
101	50871	51299	-	142	-	-	VSP based on similarity to AHIS-1 ORF 93 identified by MS; see Table 4.5
103	52448	56743	-	1431	-	-	putative tail fibre protein, based on similarity to A-HIS1 ORF 95, see Table 4.5
104	56808	57350	-	180	3e-14	-	End, top hit coliphage RB16 HNH(AP2) 6 , putative HNH endonuclease

4.3.4 Genome organisation

A-HIS1 and A-HIS2 are similar with respect to the ORFs they encode. Both genomes were divided into three main regions based on the distribution of predicted ORFs on the two DNA strands (Fig. 4.1, Fig. 4.2). The ORFs in regions I and II are reasonably strand-specific, with ORFs occurring mainly on one particular DNA strand.

However, in region III, the ORFs are distributed fairly evenly across both DNA strands. Genes involved in DNA replication, metabolism and modification are encoded mainly in region I for both phages. Virion structural proteins (VSPs) are found in regions II and III, including those identified by mass spectrometry.

4.3.5 Virion structural proteins

The blastp analysis identified only three putative VSPs in each phage: a putative minor head protein, tape measure protein and tail fibre protein. A top hit for the A-HIS1/2 putative minor head protein was the phage SPP1 gp7 family minor capsid protein, which is associated with the gp6 portal protein of SPP1 (Vinga et al., 2006). Further investigation is required to identify the A-HIS1/2 portal protein gene. Using SDS-PAGE followed by mass spectrometry, 7 and 9 ORFs were identified from A-HIS1 and A-HIS2, respectively, as VSPs (Fig. 4.3, Table 4.4). The observed protein sizes from SDS-PAGE correlated well with the predicted molecular weights.

Notably, the protein lanes of purified phage ghosts in the SDS-PAGE gels were always dominated by one particular protein band for each phage at $\sim$38 kDa, which corresponded to ORFs 73 and 82 of A-HIS1 and A-HIS2, respectively (Fig. 4.3, Table 4.4). ORFs 73 and 82 were also repeatedly identified in other protein bands on the protein gel.

It has been well-documented that the major capsid protein in siphoviruses have a much higher copy number than the major tail protein. Specifically, coliphage λ has $\sim$480 head copies to $\sim$190 tail copies with the major capsid making up $\sim$57 % of the total protein in a λ phage compared to the tail which makes up $\sim$19 % (Buchwald et al., 1970). For T5, Zweig and Cummings reported 65 % for the head and 17 % for the tail as a percentage of total protein in a T5 phage (Zweig and Cummings, 1973). This led us to conclude that these ORFs are the putative major capsid of phages A-HIS1 and A-HIS2, respectively.

A-HIS1 ORF 95 has been annotated as a putative tail fibre protein based on the blastp analysis which revealed similarity to an immunoglobulin I-set domain protein and a fibronectin type III protein. Recently, Fraser et al. found that Ig-like domains are present in many dsDNA phage structural proteins, and that these protein domains are likely to be on the outside of

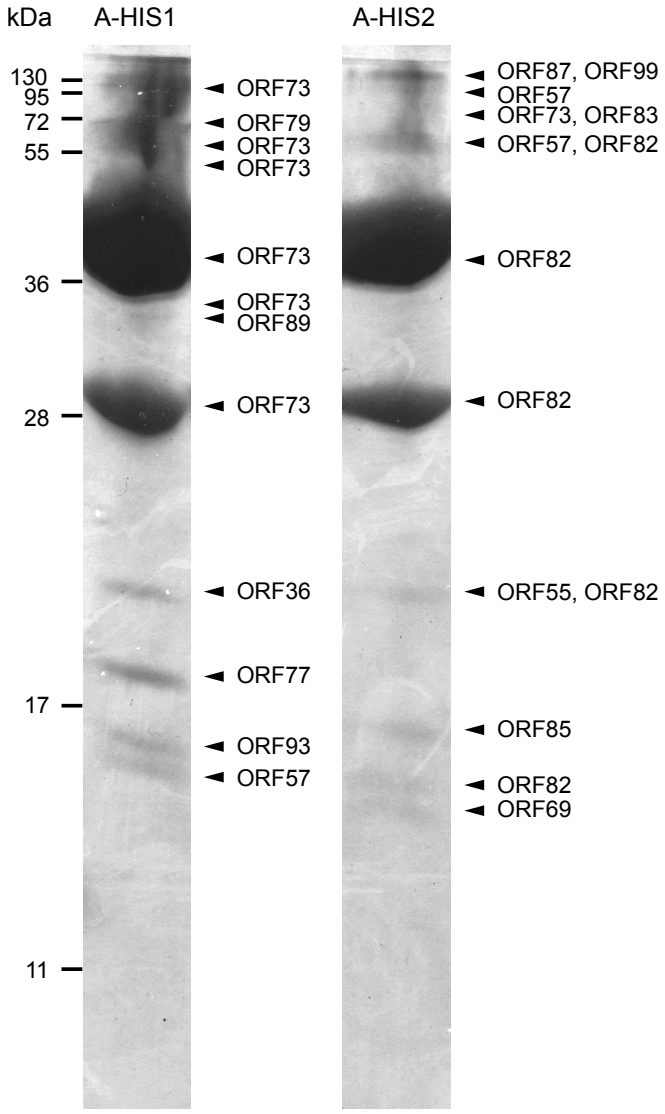


Fig. 4.3: Identification of phage structural proteins. Coomassie-stained SDS-PAGE gel of purified proteins from A-HIS1 and A-HIS2 phage ghosts. Arrows indicate the position of bands removed from the gel and the corresponding ORFs which were recognised after amino acid sequence identification.

Table 4.4: ORFs identified by mass spectrometry.

Phage	ORF	Observed MW (kDa)	Predicted MW (Da)	Putative function
A-HIS1	36	21	19,220	-
	57	15	14,965	-
	73	28, 35, 38, 55, 100	38,415	Major capsid
	77	18	17,623	-
	79	72	76,924	Tape measure protein
	89	34	27,469	-
	93	16	16,250	-
A-HIS2	55	21	19,072	-
	57	60, 95	75,109	-
	69	14.5	14,499	-
	73	73	50,449	-
	82	16, 21, 28, 38, 60	38,418	Major capsid
	83	73	56,185	-
	85	17	16,289	-
	87	120	110,918	Tape measure protein
	99	120	25,067	-

the phage (Fraser et al., 2006). A-HIS2 ORF 103 was similarly annotated as a putative tail fibre protein based on its similarity to A-HIS1 ORF 95 (see Table 4.5). However, no conserved domains were detected and therefore the function has been tentatively assigned. Using ACT, a repeated region was observed in the putative tail fibre protein encoded by ORF 103 of A-HIS2 which occurred only once in ORF 95 of A-HIS1. PSORTb localised the putative tail fibre proteins to the outermembrane or to be extracellular.

4.3.6 Comparative genomics

4.3.6.1 A-HIS1 vs A-HIS2

When the genomes of phages A-HIS1 and A-HIS2 were compared to each other a high degree of synteny was observed and in general, the relative positions of genes are conserved. The amino acid sequences of predicted ORFs from each phage were compared against each other by standalone

blastp to find the number of genes with significant identity between the two phages. The A-HIS1 genes were queried against the A-HIS2 genes as the database and vice versa. *Shared genes* were found by considering only those gene pairs which were top hits for each other in both sets of blastp results which could mutually identify each other.

In total, phages A-HIS1 and A-HIS2 share 48 such genes (Table 4.5). Other gene pairs, termed *similar genes* were identified by considering the lesser hits. All assignments were based on all percent identity matrix (PIM) values betweens ORFs being greater than 20 % when pairs of genes were aligned. This was observed for example for the AP2 endonucleases (Table 4.6, discussed further in § 4.3.7.2). It is worth noting that though these phages share nearly half their genes, they remain highly divergent from each other with % identity ranging from the least similar shared gene 20 % for A-HIS1 ORF 51 and A-HIS2 ORF 66 up to the most similar gene, the carboxypeptidase with 70 % identity between the two phages (A-HIS1 ORF 88 and A-HIS2 ORF 97).

Using NCBI blastp only ORF 59 from A-HIS2 had a significant hit (e-value $6e^{-5}$) to an *Acaryochloris* gene which encodes hypothetical protein AM1_2888. A-HIS1 ORF39 also had a hit to AM1_2888 but the e-value was 0.6. This lone result indicated there was no strong evidence for lateral gene transfer between phages A-HIS1 and A-HIS2 and the host strain MBIC11017.

A further three ORFs could be assigned a virion structural function by virtue of their similarity (from the analysis of shared genes, Table 4.6) to ORFs identified as VSPs by mass spectrometry. As a result virion structural function was assigned to A-HIS1 ORFs 37 and 63 which are similar to A-HIS2 ORFs 57 and 73, respectively. Similarly, this applied to A-HIS2 ORF 101, which is a homologue of A-HIS1 ORF 93. In total, the combination mass spectrometry, blastp analyses and ORF comparisons, allowed

Table 4.5: Shared genes. Summary of gene similarities between A-HIS1 and A-HIS2 based on the standalone blastp (default parameters) of each set of phage genes first as query against the other, then as the subject (i.e. database) for the other. Results where e $>$ 1e^{-3} were removed. Shared gene pairs were the top hits in both blastp searches which could mutually identify one another. % identity values taken from PIM calculated in ClustalX (1.83). * denotes ORFs with nucleotide sequences which are inverted with respect to each other. Abbreviations as in Table 4.2 and Table 4.3.

A-HIS1		A-HIS2		% identity	Putative function
ORF	Size (aa)	ORF	Size (aa)		
1	486	1	478	63	TerL
4	86	9	86	37	-
6	236	12	137	42	-
12	112	17	97	35	-
14	993	20	959	57	DNA pol γ
16	605	22	610	47	DNA pri
19	283	25	238	40	-
20	191	27	199	47	RNase T
22	57	28	63	37	-
24	209	30	210	49	-
25	59	95	58	36	-
26	108	31	141	50	End
27	269	32	251	36	DNA Mtase
28	259	33	345	42	-
30	365	35	361	52	DNA pol I-like FEN
31	265	36	235	35	-
33	263	38	243	32	-
34	73	44	73	31	-
36	169	55	168	61	VSP
37	688	57	687	66	VSP
41	55	58	50	50	-
43	83	54	83	64	-
46	149	45	133	27	-
47	61	63	76	33	-
51	267	66	276	20	-
53	77	67	95	42	-
55	539	68	278	49	min head
57	137	69	128	53	VSP
58*	176	104	180	56	End
63	460	73	443	57	VSP
64	190	74	201	36	-
65	61	76	59	30	-
66	214	77	208	43	-
70	78	79	78	31	-
71	54	81	51	31	-
73	359	82	358	60	maj head
75	173	84	174	62	-
77	156	85	146	57	VSP
79	737	87	1041	52	TMP
82	216	90	199	56	dNK
84	89	91	95	30	-
86	160	94	158	50	-
88	371	97	371	70	Carbpep
89*	260	99	239	59	VSP
91	96	42	86	53	-
93	147	101	142	28	VSP
94	376	102	378	45	-
95	955	103	1431	43	tail fibre

Table 4.6: Similar genes. Similar gene pairs listed are those which could also mutually identify one another, but were not mutual top hits in the blastp searches. +, with the exception of A-HIS2 ORF 36. Details as in Table 4.5.

A-HIS1		A-HIS2		% identity	Putative function
ORF	Size (aa)	ORF	Size (aa)		
13	215	36/104	235/180	12/52	End+
17	182			14/48	
58	176	36	235	15	
41	55	4	60	29	-
43	83	42*	86	46	-
76	127	45	133	24	-
91	96	54*	83	44	-

23 ORFs of each phage to be assigned a putative function (Table 4.2 and Table 4.3), i.e. 24 % (A-HIS1) and 22 % (A-HIS2).

4.3.6.2 A-HIS1/2 vs other siphoviruses

Phage taxonomy has aroused much discussion in recent years and there has been growing support for a comparative genomics led-approach to understanding *Caudovirales* diversity (Nelson, 2004). However, new methodologies for taxonomy are being highlighted along with the dilemmas of existing methods, since there does not appear to be a unifying taxonomic method (Proux et al., 2002). In fact, comparative analyses of the *Acaryochloris* phage genomes showed they have a unique composition and organisation compared to other siphoviruses. Seguritan et al. and Brüssow and Desiere have shown that other siphoviruses appear to have some conservation of gene order which is not the same as that found in the *Acaryochloris* phages (Brüssow and Desiere, 2001; Seguritan et al., 2003). Another feature of other siphoviruses from these genome comparisons is that each has a cluster of genes comprising most or all of the following genes: a large terminase, a small terminase, a portal protein, a protease and a major head protein. These genes occur very close or next to each other in more or less the same order in the examples presented.

The equivalent genes in the *Acaryochloris* phages are much more separated

by unknown genes (in this case, the protease identified is the carboxypeptidase, although no portal protein has been identified). A-HIS1 and A-HIS2 also contain novel genes found in no other phages like an RNase T gene. In light of all these observations it would be premature to place them into an existing *Siphoviridae* genus.

4.3.6.3 A-HIS1/2 vs strain MBIC11017

The genome of strain MBIC11017 was recently sequenced and there are a number of phage-related genes in the *A. marina* MBIC11017 genome, mainly consisting of tail-related proteins, a phage lysozyme and many integrases which are commonly associated with temperate phage. However, from a blastp analysis it seems that none of these genes are related to genes in the phages A-HIS1 and A-HIS2. Moreover, unlike the situation with *Synechococcus* and *Prochlorococcus* phages, no photosynthesis-related genes were found in phages A-HIS1 and A-HIS2. This property may extend to other *Acaryochloris* phages and may be because they do not infect a host which is found in high-light conditions where genes such as *psb*A are thought to provide increased fitness under high irradiance levels (Bragg and Chisholm, 2008; Hellweger, 2009).

4.3.7 Nucleic acid metabolism and modification

A-HIS1 and A-HIS2 encode several different proteins involved in nucleic acid metabolism and modification. A-HIS1 also encodes a DNA methylase (ORF 38) and A-HIS2 ORF 34 exhibits similarity to the N-terminal half of ECF-type RNA polymerase sigma-24 factors, which had best match to *Ruegeria* sp. TM1040 using NCBI BLAST (Table 4.3), a representative of a group of abundant marine α-proteobacteria (Gonzalez and Moran, 1997).
A flap endonuclease (FEN), AP2 endonucleases, mtDNA polymerase (pol) γ and RNase T were also identified in both phage genomes. These genes

are considered in more detail individually. In particular, the latter two have never been found in a phage genome before, let alone a cyanobacterial genome and are therefore considered together in a case study in more detail in § 4.3.8.

4.3.7.1 Flap endonucleases

Amongst the *Acaryochloris* phage-encoded proteins involved in nucleic acid metabolism and modification were A-HIS1 ORF 30 and A-HIS2 ORF 35, which encode a protein similar to bacterial DNA polymerase I sequences identified by NCBI BLAST. However, DNA polymerase I sequences contain both a C-terminal DNA polymerase A domain and a N-terminal 5'-3' exonuclease domain, whereas these phage ORFs only contain the latter domain. These proteins belong to a group of proteins called flap endonucleases or FENs (Harrington and Lieber, 1994) and are putative DNA polymerase I 5'-3' exonuclease paralogues (Allen et al., 2009).

These enzymatic proteins have been found in viruses and all domains of life and are essential in DNA replication. Recently, it has been shown that FEN paralogues can be separated into two classes based on their amino acid composition; those which have all the amino acids to bind two metal ions, e.g. the *Staphylococcus aureus* 5'-3' exonuclease or *Sa*FEN, and those which only bind one, e.g. the *Escherichia coli* exonuclease IX or *Ec*ExoIX (Allen et al., 2009). In fact, the second metal site binding properties of the DNA polymerase I 5'-3' exonuclease paralogues are largely based on the presence of three aspartate residues (Fig. 4.4, (Allen et al., 2009)), which exonuclease IX sequences lack. In particular, Allen et al. did not include the *Shewanella* sp. or *Colwellia* sp. tree nodes in their *Sa*FEN/*Ec*ExoIX groupings in their phylogeny, since they fell in between the two types of FEN paralogue and it can be observed that these sequences lack two of the three aspartate residues Fig. 4.4. Clearly, the phage FEN homologues

contain all three aspartate residues and therefore are similar to the 5'-3' exonuclease FENs.

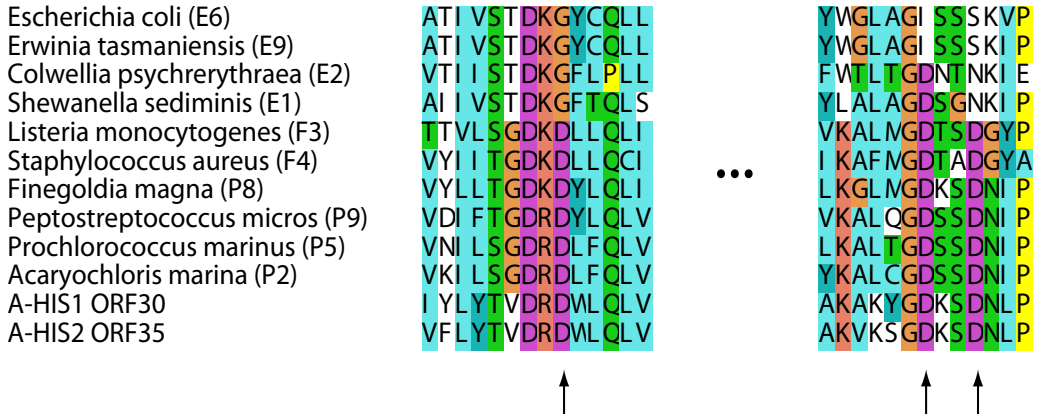


Fig. 4.4: Second metal site residues of FEN paralogues. Alignment of a selection of exonuclease IX proteins (E), DNA polymerase I 5'-3' exonuclease paralogues (F) and N-terminal DNA polymerase I 5'-3' exonuclease domains (P) with *Acaryochloris* phage FEN homologues. Positions responsible for second metal site binding indicated by arrows (see (Allen et al., 2009)). Assignments in parentheses correspond to sequences shown in Fig. 4.5; details in Appendix A Table A.3.

In order to aid the annotation of the phage FEN homologues, a phylogenetic analysis was computed based on that of Allen et al. (Allen et al., 2009). First the sequences used by Allen et al. were identified in NCBI BLAST. To this, the top BLAST sequence matches to the phage FEN homologues were added. As mentioned previously, these were bacterial DNA polymerase I sequences which contain a DNA polymerase A domain as well as the 5'-3' exonuclease domain. The host *A. marina* also contains a DNA polymerase I protein and therefore the N-terminal domain of this protein was also included. To form the alignment, the sequences used in Allen et al. were aligned with the BLAST matches and the *A. marina* sequence, followed by a profile alignment to the phage FEN homologues. The alignment was then truncated by removing the C-terminal portion of the alignment containing the DNA polymerase A domain of the DNA polymerase I sequences.

The phylogenetic tree computed from the new alignment showed the same general topology to that of Allen et al. of FEN-like proteins Fig. 4.5. It can be clearly seen that the DNA polymerase I 5'-3' exonuclease sequences

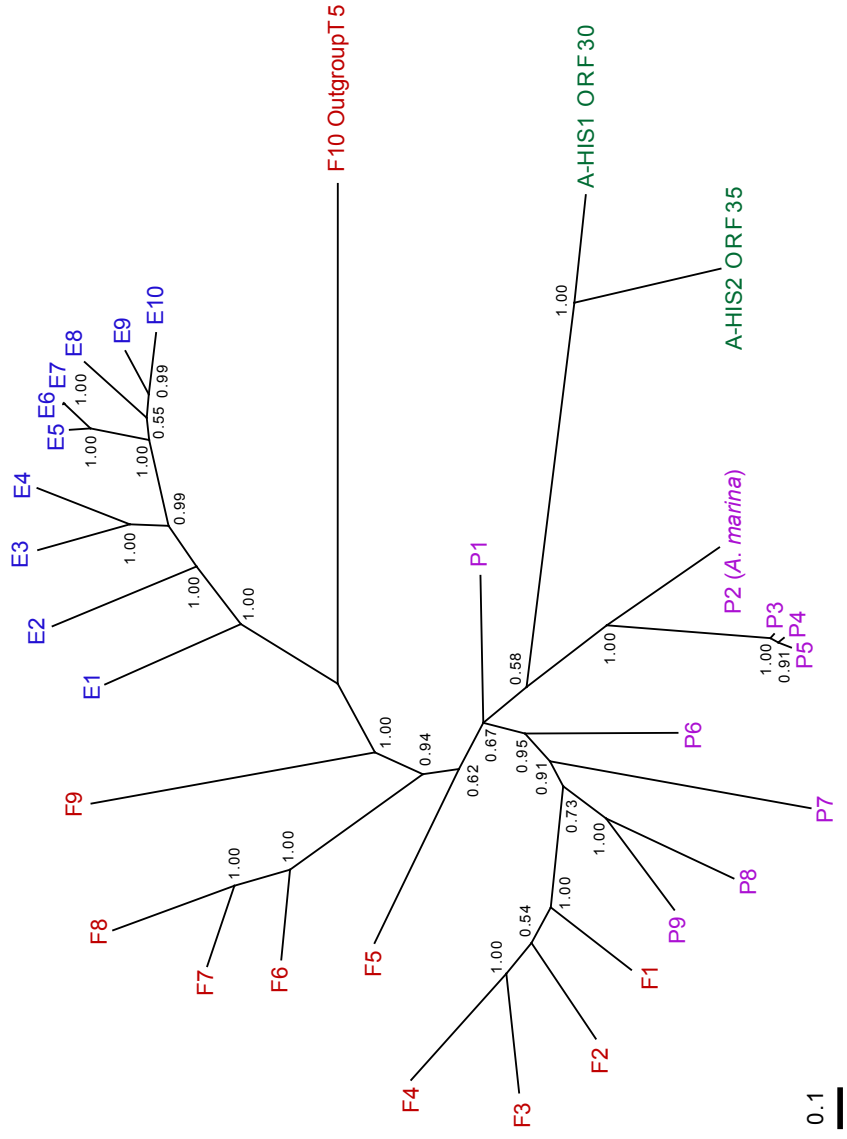


Fig. 4.5: Phylogeny of exonuclease paralogues. Bayesian phylogenetic tree of exonuclease IX proteins, DNA polymerase I 5′-3′ exonuclease paralogues, N-terminal DNA polymerase I 5′-3′ exonuclease domains and *Acaryochloris* phage FEN homologues. All sequences are bacterial except for the *Acaryochloris* phage sequences (green) and that of the phage T5 FEN (outgroup). In particular, P3-P5 are cyanobacterial sequences (*P. marinus*). E = Exonuclease IX (blue), F = FEN containing 5′-3′ exonuclease domain (red), and P = DNA polymerase I sequences (purple). See Table A.3 for sequence details. Scale bar units: amino acid substitutions per site.

fall with the FEN paralogues which contain only the 5'-3' exonuclease domain. Interestingly, the *Acaryochloris* phage exonucleases grouped with the cyanobacterial DNA polymerase I domains of *A. marina* (sequence P2) and *P. marinus* (sequences P3, P4 and P5). However, these phage proteins did not group with either of the *Sa*FEN-/*Ec*ExoIX-type proteins. Therefore, A-HIS1 ORF 30 and A-HIS2 ORF 35 were annotated as DNA polymerase I-like FEN homologues, also called 5'-3' exonucleases.

4.3.7.2 AP2 endonucleases

A-HIS1 and A-HIS2 were predicted to contain 4 and 2 ORFs which possess an AP2 domain from the conserved domains or blastp analyses (Table 4.2, Table 4.3). These are ORFs 13, 17, 26 and 58, and ORFs 31 and 104 (referred to here as the End genes), respectively. The AP2 domain is associated with a family of transcription factors in plants, though has been found present in phage, bacterial and eukaryotic genes (Magnani et al., 2004; Lee et al., 2007). The proteins encoded by these genes with AP2 domain(s) have been predicted to be HNH endonucleases or homing endonucleases, which are mobile genetic elements found in all biological kingdoms including mitochondria and chloroplasts (Shub et al., 1994; Gimble, 2000). The *Acaryochloris* phage AP2 End genes were aligned using an existing alignment by Magnani et al. of nonplant proteins which contain AP2 domains (Fig. 4.6). From the alignment, it was evident that while all of the *Acaryochloris* phage End genes possessed an AP2 domain similar to that of the other sequences in the alignment, A-HIS1 ORF 26 and A-HIS2 ORF 31 appear to lack most if not all of the histidine/asparagine (H/N) features of an HNH domain.

To compare the AP2 endonucleases to each other, a sequence alignment with reduced gap penalties from default § 4.2.3, was performed and the PIM was analysed. A-HIS1 ORFs 13, 17, 58 and A-HIS2 ORF 104 had a

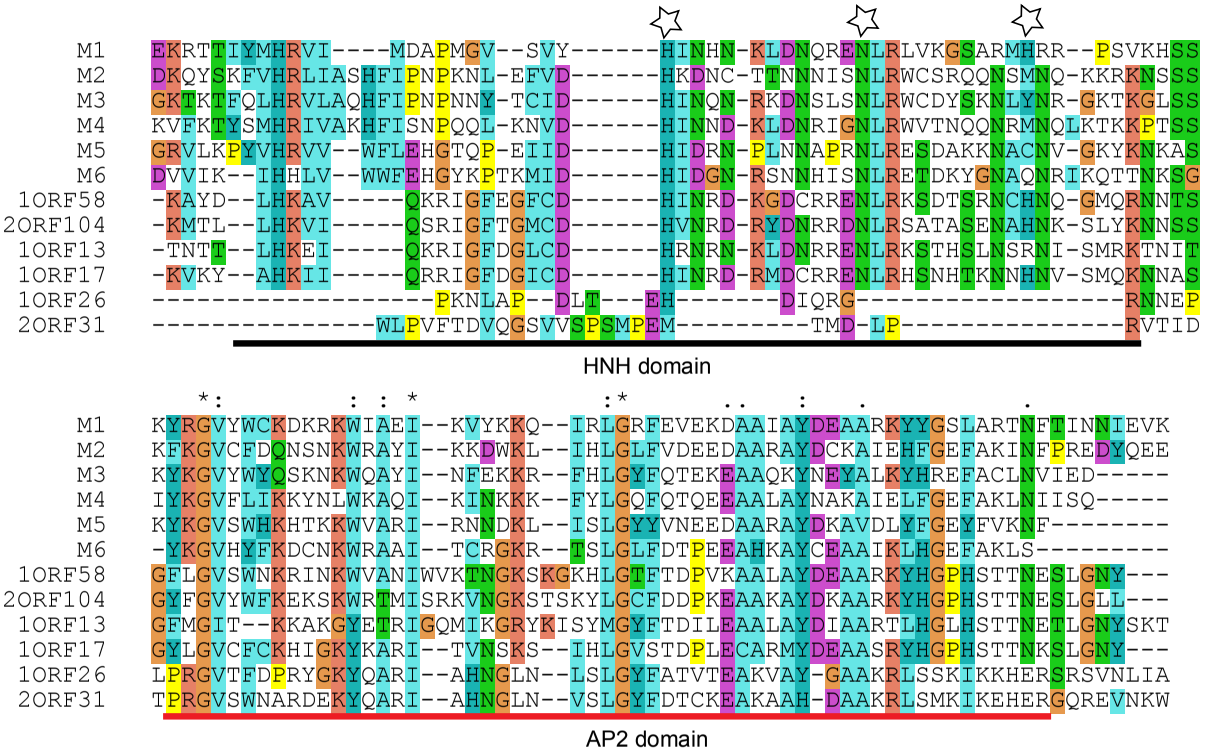


Fig. 4.6: Alignment of putative HNH endonucleases. ClustalX alignment of *Acaryochloris* End genes to that of non-plant proteins which contain AP2 domains by (Magnani et al., 2004) denoted by the sequences M1-M6: M1, *T. erythraeum*, M2-M4, *T. thermophila* proteins, M5, Enterobacteria phage RB49, and phage Felix 01. The domains are marked by the black and red lines. The open stars denote the residue positions which give rise to the HNH nomenclature.

good degree of similarity with PIM values > 45 % between each other after alignment (Table 4.7). A-HIS1 ORF 26 and A-HIS2 ORF 31 were similar to each other sharing 62 % identity, while their % identity to the other AP2 endonucleases was less than 24 %, which correlates with the observations from the alignment of these genes (Fig. 4.6).

Table 4.7: AP2 endonuclease sequence comparison. Percent identity matrix comparing *Acaryochloris* phage AP2 endonucleases. Pairwise/multiple gap opening and extension penalties 2 and 0.1, respectively. ORFs 13, 17, 26 and 58 are from A-HIS1 and ORFs 31 and 104 are from A-HIS2.

ORF	17	58	104	13	26	31
17	100	58	49	46	22	19
58	58	100	56	52	16	12
104	49	56	100	50	24	19
13	46	52	50	100	18	14
26	22	16	24	18	100	62
31	19	12	19	14	62	100

Specifically, homing describes the lateral transfer of an intron (or intein) in a gene to a homologous allele of that gene which does not contain it and the homing process is catalysed by homing endonucleases, which are encoded within the intervening sequence (Chevalier and Stoddard, 2001). To check whether the End genes encode inteins, they were checked against the intein database, InBase (Perler, 2002), however, no matches were found.

In 2004, Mehta et al. classified the available HNH family of protein sequences into eight subsets based on their combination of different features in the HNH domain (Mehta et al., 2004). The second subset in particular, contained mainly phage proteins, a subset of which are associate with the AP2 domain. Using the classifications of Mehta et al. none of the *Acaryochloris* phage End genes have a GG motif, and therefore do not fall into subsets 1, 3, 5, 6 or 7. They also are not members of subsets 4 and 8 since they do not contain cysteine dyads. Subset two is defined by HNN in the HNH-element, where there is an N instead of an H in the motif. Only A-HIS1 ORFs 13 satisfies the features of this subset. In A-HIS1 ORF 17 and A-HIS2 104 the final N is an H. In A-HIS1 ORF 58, there is also a final H

residue, however, this is preceded by a single cysteine residue, where there is usually a cysteine dyad. A-HIS1 ORF 26 appears to only retain a single histidine residue in the HNH-element (manually aligned, see open star positions in Fig. 4.6), whereas A-HIS2 ORF 31 retains no histidine or asparagine residues which denote the HNH-element. However, A-HIS1 ORF 26 and A-HIS2 ORF 31 do contain the AP2 domain, and therefore, these HNH endonucleases appear to have degenerated.

Mehta et al. noted that most of the annotated proteins belonging to the second subset of HNH family proteins are intron-encoded site specific endonucleases, and therefore, proteins which had not yet been annotated in this subset may have a similar function. This suggests that A-HIS1 ORF 13 and the other AP2 endonucleases of A-HIS1 and A-HIS2 are likely to be intron-encoded, which has also been suggested of the HNN-AP2 genes of *Xanthomonas oryzae* phages (Lee et al., 2007). In considering the End genes as introns, the two neighbouring ORFs either side of each End gene were assessed to examine if they were likely to be part of the same gene, but interrupted by their respective End gene.

On analysis of End gene A-HIS1 ORF 13, the surrounding ORFs 12 and 14 were amalgamated by addending A-HIS1 ORF 12 to the C-terminus of A-HIS1 ORF 14. The amalgamated ORF was then submitted to blastp, and alignments were assessed which indicated that there was possible sequence similarity between A-HIS1 ORF 12 and the C-terminus sequence of some blast hits. Blast sequences were subsequently aligned with A-HIS1 ORF 14, the A-HIS2 homologue; A-HIS2 ORF 20 and the amalgamated sequence denoted 14_12 (see Fig. 4.7). The alignment indicated that there was a region which was fairly well conserved between the DNA polymerase γ sequences from the blastp results and the addended A-HIS1 ORF 12, which was not assigned a putative function previously based on poor blastp hits. The same analysis on the other End genes was inconclusive, since unlike

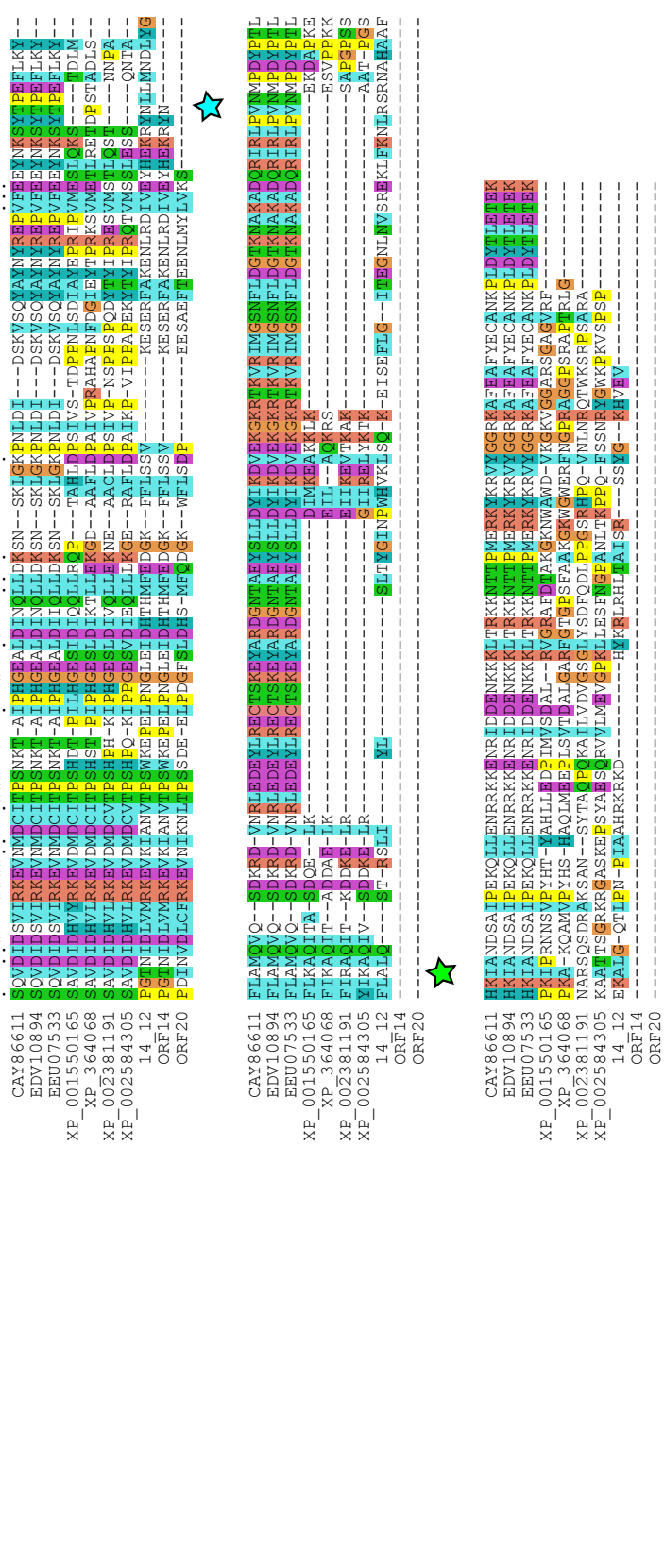


Fig. 4.7: Alignment of DNA polymerase γ C-terminus. ClustalX alignment of *Acaryochloris* phage mtDNA polymerase γ sequences: A-HIS1 ORF 14 and A-HIS2 ORF 20 with blast hits of amalgamated ORF (14_12) produced by addending A-HIS1 ORF 12 to C-terminus of A-HIS1 ORF 14. Blue star denotes the A-HIS1 ORF 14 C-terminus (A-HIS1 ORF 12 begins LLMND). Green star denotes region of high sequence conservation. Blast hits are denoted by their accession numbers. *S. cerevisiae* proteins: CAY86611, Mip1p str. EC1118; EDV10894, mtDNA polymerase catalytic subunit (str. RM11-1a) and EEU07533, Mip1p str. JAY291; XP_001550165, hypothetical protein BC1G_11008 *B. fuckeliana* B05.10; XP_364068, hypothetical protein MGG_08913 *M. grisea* 70-15; XP_002381191, DNA polymerase γ *A. flavus* NRRL3357 and XP_002584305, DNA polymerase γ *U. reesii* 1704.

A-HIS1 ORF 14 which encodes mtDNA polymerase γ, the blastp results available to perform a comparison were limited. In conclusion, the observations made here suggest that the End genes are intron-encoded. Further experiments are required to confirm this hypothesis.

4.3.8 DNA polγ and RNase T

4.3.8.1 DNA polγ: background and hypotheses

DNA polymerase is an enzyme that catalyses the polymerization of deoxyribonucleotides into a DNA strand and is therefore essential to all known forms of life on earth. The gene which encodes DNA polymerase is a useful tool to interpret evolutionary relationships between different organisms (Braithwaite and Ito, 1993; Filee et al., 2002; Labonté et al., 2009). There are currently seven families of DNA polymerase - A, B, C, D, X, Y and RT (Filee et al., 2002; Harada et al., 2005). DNA polymerase γ is a high fidelity family A DNA polymerase found in mitochondria (Wang, 1996), which is responsible for the replication of mitochondrial DNA (mtDNA) in the mitochondria of eukaryotes (Bolden et al., 1977). The origin of mtDNA polymerase γ in mitochondria is unknown because the bacterium which is thought to have led to the modern day mitochondria of eukaryotes, has never been identified.

Several proteins involved in mitochondrial DNA replication and gene expression are related to corresponding phage proteins, namely the DNA-directed RNA polymerase, the helicase TWINKLE (Shutt and Gray, 2006b,a; Falkenberg et al., 2007) and from these also several proteins evolved which control these processes in plant chloroplasts (Hess and Borner, 1999). However, it has been one of the biggest surprises that even core components of the mtDNA polymerase, in particular the catalytic γ subunit, are similar to proteins known from the T-odd lineage of bacteriophages (Shutt and Gray, 2006a; Falkenberg et al., 2007). It has been hypothesised that

mitochondria may have inherited DNA polymerase γ by phage-mediated non-orthologous displacement (Koonin et al., 1996) based on phylogenetic studies and T3/T7-like cryptic prophages which were found in several proteobacterial genomes (Filee et al., 2002; Filee and Forterre, 2005), since its origins remained elusive as DNA polymerase γ is not found in any known bacteria. Until now this has remained speculative as no such sequences have actually been found in a phage but this new data provided by these phages suggests there is some support for this hypothesis.

However, it could be that either the ancestors of these bacteriophages infected a protomitochondrion-like bacterium, which may be extinct or alternatively they still infect such a bacterium in addition to their cyanobacterial hosts. To examine these alternative hypotheses, CAMERA (Seshadri et al., 2007), a metagenomics data repository, was searched with the phage mtDNA polymerase to see if it was possible to find homologues which could be included in phylogenetic analyses. In addition, it was of interest to investigate whether the neighbouring genes to the homologues on an assembly could help identify what organism they came from.

4.3.8.2 DNA polγ: sequence comparison

Blastp against the NCBI nr database showed the two *Acaryochloris* phages possess an ORF coding for a mtDNA polymerase, similar to that of fungi and metazoans, which has never been found in a phage genome before. Comparative analyses showed the two phage DNA polymerase γ proteins share 57 % identity (Table 4.5). In comparison, the most recent (May 2009) top fungal and *Homo sapiens* sequence from each set of BLAST results for the two phage DNA polymerases share 28–31 % identity with the phage DNA polymerase sequences (Table 4.8). A decrease in the gap opening and extension penalties to 2 and 0.1 from the default values, respectively, raised the PIM values by 1–4%.

Table 4.8: DNA polymerase γ sequence comparison. Percent identity matrix comparing *Acaryochloris* phage genes with NCBI sequences using Clustal X. F = fungi, H = human, G = phage Gamma. Sequence details include NCBI accession numbers.

F1	F2	H1	H2	G1	G2	%	Sequence details
100	62	41	42	31	31	F1	XP_001550165: *Botryotinia fuckeliana* B05.10
62	100	42	42	30	31	F2	EEH49594: *Paracoccidioides brasiliensis* Pb18
41	42	100	100	28	30	H1	BAF83970: *Homo sapiens*
42	42	100	100	28	30	H2	NP_002684: *Homo sapiens*
31	30	28	28	100	57	G1	A-HIS1 ORF 14
31	31	30	30	57	100	G2	A-HIS2 ORF 20

4.3.8.3 DNA polγ: phylogenetics

A phylogenetic analysis of the phage mtDNA polymerases was performed, using the DNA polymerase A alignment by Filée et al. as a framework (Filee et al., 2002), which contained the sequences of the family A DNA polymerase A domain (see § 4.2.6 for details).

To select sequences to add to the alignment, it was noted that A-HIS1 and A-HIS2 also encode a DNA polymerase I-like FEN of bacterial origin (A-HIS1 ORF 30 and A-HIS2 ORF 35), which has blastp hits to DNA polymerase I, a family A DNA polymerase (§ 4.3.7.1). Therefore it was of interest to include these blastp hits in the phylogenetic analysis, although the phage FENs themselves would be excluded given they do not contain a DNA polymerase A domain. In particular, among the bacterial hits were cyanobacterial DNA polymerase I sequences, which do not appear in the alignment by Filée et al. and specifically that of *A. marina* and the inclusion of these sequences would allow their comparison to the mtDNA polymerases. Other family A DNA polymerase sequences which were added to this alignment included the plant *Oryza sativa* as an outgroup, those of the *Caudovirales* families - *Myoviridae*, *Siphoviridae* and *Podoviridae* and DNA polymerase γ sequences from fungi (from the top

blastp hits to the phage mtDNA polymerase) and of course the *A. marina* phage mtDNA polymerase sequences. Finally, sequences were retrieved from the metagenomic ORFs peptides dataset of the CAMERA database (Seshadri et al., 2007) by blastp of the phage mtDNA polymerases. In total, 12 assemblies were found which contained phage mtDNA polymerase-like sequences, which either had a partial or complete DNA polymerase A conserved domain. Of the 12 phage mtDNA polymerase-like sequences, there were three sequences (A05ORF9, A06ORF10 and A08ORF1) which could be completely aligned and one (A02ORF1) which had almost complete alignment since the sequence was not long enough (Appendix A, Fig. A.2). Four main clades were identified in the phylogenetic tree shown in Fig. 4.8; a mtDNA polymerase γ clade, a *Siphoviridae* clade, a *Podoviridae* clade and finally a bacterial clade which encompasses bacterial, cyanobacterial and Mus308 sequences (Harris et al., 1996). The *Myoviridae* sequences did not form a specific clade: the myovirus *Bacillus* phage SPO1 fell deeply in the bacterial clade, which suggests it may be of bacterial origin, the *Burkholderia* phage BcepNY3 did not fall into a clade and the coliphage phiEcoM-GJ1 rooted with the *Podoviridae* clade. Intriguingly, the two *Acaryochloris* phage mtDNA polymerases branched deeply within the mtDNA polymerase γ clade, forming a sister clade to the fungal and metazoan DNA polymerase γ proteins. The samples identified as DNA polymerase γ sequences from the CAMERA database were also clearly grouped within the same clade.

The observation that *Acaryochloris* phage sequences fall with the mtDNA polymerase γ sequences provided the first evidence that this gene may have originated in a bacteriophage and supported the hypothesis Filée et al. proposed. An examination of the identity of the other predicted ORFs in the CAMERA assemblies on which the phage mtDNA polymerase-like genes were found, was then used to see whether the origin of each assembly could

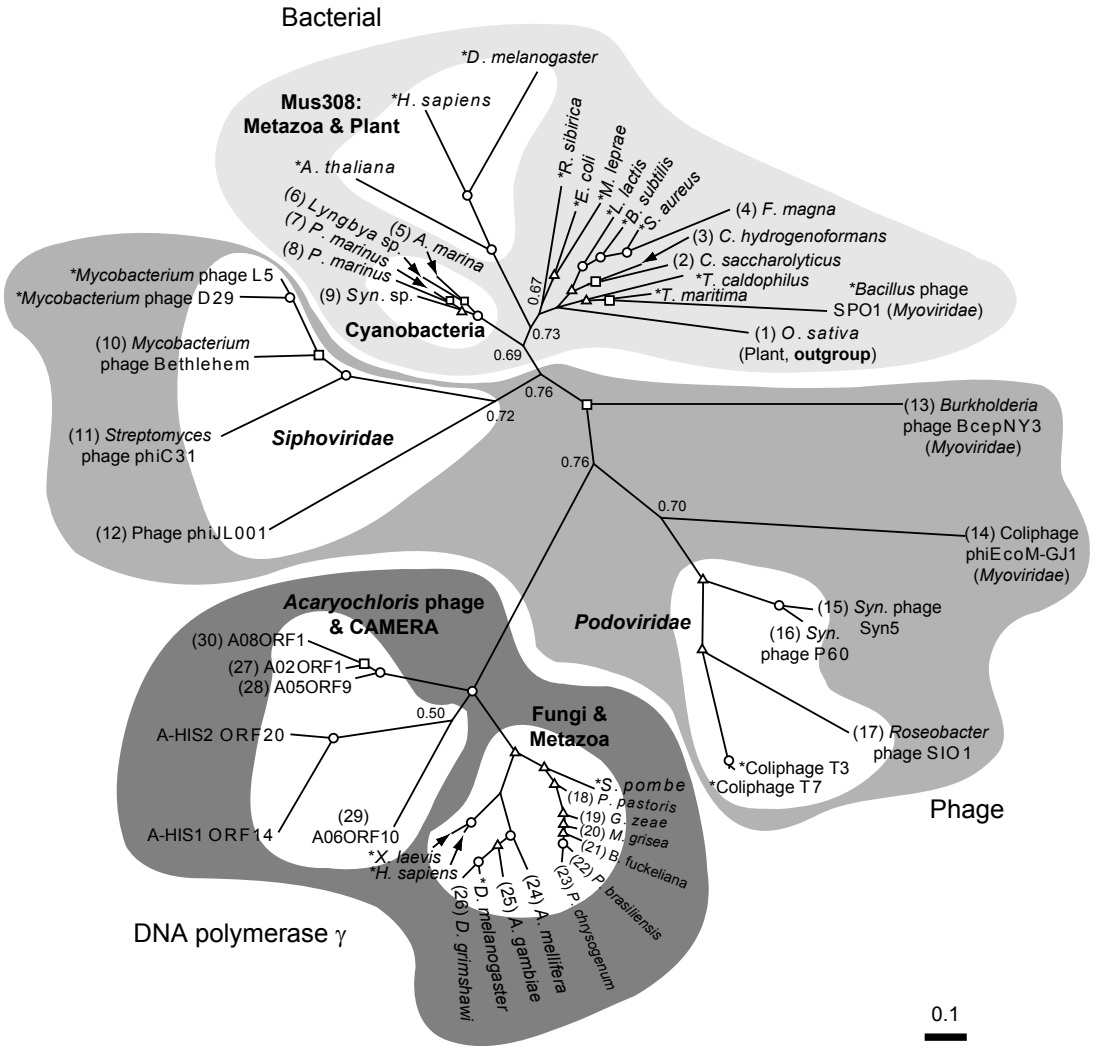


Fig. 4.8: Phylogeny of family A DNA polymerases. Bayesian phylogenetic tree constructed in MrBayes; support values are either explicit or denoted by node symbols; circle (1.00), triangle (>0.90), square (>0.8). Nodes labelled AxORFy refer to sequences from CAMERA. Numbers in parentheses refer to sequence details provided in Fig. A.2 (Appendix A). Sequences (28) and (29) may be of bacterial or phage origin, from Punta Cormorant (hypersaline lagoon), Ecuador (Table 4.9 and Table 4.10). Sequences of unknown origin (27) and (30) from Lake Gatun (freshwater), Panama and Rangirora Atoll (coral reef atoll), French Polynesia, respectively. * denotes sequences from alignment by Filée et al. (Filee et al., 2002). Tree nodes have been labelled with support values. Scale bar units: amino acid substitutions per site. *Syn.* = *Synechococcus*.

be identified (§ 4.3.8.4).

4.3.8.4 DNA polγ: CAMERA assemblies

Of the 12 assemblies which contained an *A. marina* phage mtDNA polymerase γ-like sequence, four contained other predicted genes with significant BLAST hits besides the mtDNA polymerase, which may be of bacterial or phage origin: A01ORF4, A04ORF3, A05ORF9 and A06ORF10 (Table 4.9 and Table 4.10). It was noted in assembly A05 that ORF A05ORF8 had a blastp hit to S-PM2p158 which encodes a rare lipoprotein A, which has not been well-characterised in the literature. S-PM2 is the only phage which encodes this gene, which is predominantly found in bacteria and eukaryotes, which suggests that A05ORF8 may not be of phage origin. In addition, the top BLAST hits for A05ORF8 were bacterial, which suggests this gene and thus the assembly it belongs to, may be of bacterial origin. Assemblies A01 and A04 may be of phage origin based on the ORFs identified. For now, the four assemblies A01, A04, A05 and A06 are assigned putative bacterial or phage origin until further evidence might suggest otherwise. Unlike assemblies A05 and A06, A01 and A04 contained mtDNA polymerase γ sequences which did not contain the DNA pol A domain and so were not included in the phylogenetic tree (Fig. 4.8). An analysis of the accompanying assembly metadata indicated the size fraction that contained the CAMERA mtDNA polymerase sequences was the 0.1–0.8 μm filter fraction which was reported to be the sample that concentrated mostly bacteria and archaea (Rusch et al., 2007; Yooseph et al., 2007). However, it cannot be ruled out that this fraction may have contained viral sequences.

This analysis suggests that these environmental fragments of DNA may have originated from either phages or extant marine bacteria which are descendants of protomitochondrion-like bacteria. Indeed, if these environmental sequences came from an extant lineage of bacteria, the phylogenetic

Table 4.9: ORFs from CAMERA assemblies. Blastp analysis of ORFs from CAMERA environmental assemblies. DNA polymerase gamma ORF numbers highlighted in bold. AxORFy denotes Assembly x ORF y (the assembly numbering was arbitrary).

Assembly	Assembly ID	Location (region, country, habitat)	ORF	Comment	Putative origin
A01	1096627020476	Lake Gatun (Panama Canal, Panama, freshwater)	A01ORF1	similar to hypothetical protein S-PM2p148/gp191 phage syn9	Bacterial or phage
			A01ORF2	-	
			A01ORF3	hypothetical protein; various hits including bacterial (specifically proteobacterial), archaeal, phage, eukaryota and phycodnaviridae	
			A01ORF4	DNA polymerase gamma (no DNA pol A domain)	
A02	1096627024361		A02ORF1	DNA polymerase gamma (complete DNA pol A domain, possibly incomplete gene)	Unknown
A03	1101667175471		A03ORF1	DNA polymerase gamma (partial DNA pol A domain)	Unknown
A04	1096627138623	Punta Cormorant, Hypersaline Lagoon, Floreana Island (Galapagos Islands, Ecuador, hypersaline)	A04ORF1	putative high light inducible protein, cyanobacterial hits	Bacterial or phage
			A04ORF2	similar to hypothetical protein S-PM2p148/gp191 phage syn9	
			A04ORF3	DNA polymerase gamma (no DNA pol A domain)	
A05	1096627359208		A05ORF1	-	Bacterial or phage
			A05ORF2	-	
			A05ORF3	-	
			A05ORF4	-	
			A05ORF5	-	
			A05ORF6	-	
			A05ORF7	DPS (DNA Protecting protein under Starved conditions) domain, superfamily of ferritin-like diiron-carboxylate proteins	
			A05ORF8	rare lipoprotein A (DPBB_1 superfamily); cyanobacterial and other bacterial hits, also hit cyanophage S-PM2p158	
			A05ORF9	DNA polymerase gamma (complete DNA pol A domain, possibly incomplete gene)	

Table 4.10: ORFs from CAMERA assemblies. Table 4.9 continued.

Assembly	Assembly ID	Location (region, country, habitat)	ORF	Comment	Putative origin
A06	1096627384653	Punta Cormorant, Hypersaline Lagoon, Floreana Island (Galapagos Islands, Ecuador, hypersaline)	A06ORF1	top hit serine/threonine protein phosphatase (Thermotoga neopolitana)	Bacterial or phage
			A06ORF2	-	
			A06ORF3	Partial P-loop NTPase superfamily domain, top hits bacterial, less significant hits -phage	
			A06ORF4	Partial PP2Ac superfamily domain - bis(5'-nucleosyl)-tetraphosphatase PrpE, bacterial	
			A06ORF5	-	
			A06ORF6	eukaryota hits including methyl-CpG binding domain proteins, no putative conserved domain detected	
			A06ORF7	-	
			A06ORF8	-	
			A06ORF9	-	
			A06ORF10	DNA polymerase gamma (complete DNA pol A domain, possibly complete gene)	
A07	1096627182099	Rangirora Atoll (Polynesia Archipelagos, Fr. Polynesia, Coral Reef Atoll)	A07ORF1	-	Unknown
			A07ORF2	-	
			A07ORF3	-	
			A07ORF4	-	
			A07ORF5	DNA polymerase gamma (partial DNA pol A domain)	
A08	1096627376200		A08ORF1	DNA polymerase gamma (complete DNA pol A domain, possibly incomplete gene)	Unknown
A09	1096628021027		A09ORF1	DNA polymerase gamma (partial DNA pol A domain)	Unknown
			A09ORF2	-	
			A09ORF3	-	
A10	1096628023999		A10ORF1	Fungal DNA polymerase gamma hits containing DnaQ like exonuclease superfamily domain	Unknown
			A10ORF2	DNA polymerase gamma (partial DNA pol A domain)	
A11	1101668141779	Chesapeake Bay, MD (North American East Coast, USA, estuary)	A11ORF1	DNA polymerase gamma (partial DNA pol A domain)	Unknown
A12	1101668711231	Cabo Marshall, Isabella Island (Galapagos Islands, Ecuador, coastal)	A12ORF1	DNA polymerase gamma (partial DNA pol A domain)	Unknown

tree in Fig. 4.8 shows that they do not fall within the bacterial clade and therefore represent a new phylum of bacteria according to this gene.

4.3.8.5 RNase T: background and hypotheses

RNase T[5] belongs to the DEDDh family of proteins (Zuo and Deutscher, 2001). RNase T was first discovered in 1984 by Deutscher et al. who showed that RNase T removes adenosine monophosphate (AMP) from the 3' CCA terminus of specific tRNAs in the tRNA end-turnover process (Deutscher et al., 1984, 1985; Deutscher and Marlor, 1985). However, the physiological function of this process remains unknown. RNase T has been reported to have many functions, amongst them it may play a role in ribosomal RNA maturation (Li and Deutscher, 1995; Li et al., 1999). As observed by Deutscher (Deutscher, 1973), RNase T may act as a regulator of protein synthesis by controlling which tRNAs have their AMP cleaved as suggested by a model proposed by Stent (Stent, 1964), which may therefore specifically direct the production of proteins necessary for phage replication. However, the introduction of the RNase T gene from the *Acaryochloris* phages may serve a more obvious purpose in that a pool of AMP would be released from idle tRNAs, hence triggering tRNA degradation essentially recycling the raw material of the host as needed. Notably, RNase T has never been found in a phage before, although other ribonucleases have; a number of phages including T4 and S-PM2 encode RNase H; a ribonuclease which removes the RNA primers within Okazaki fragments (Hollingsworth and Nossal, 1991) and RNase II has also been found in lactococcal phage Q54 (Fortier et al., 2006).

Alternatively, RNase T activity may be part of an anti-CRISPR activity (pers. comms. Wolfgang Hess, see § 4.3.10.5 on CRISPRs identified in the genome of *A. marina*). It is well known that (eukaryotic) viruses have

[5] *rnt* is the associated gene name.

evolved strategies to avoid host defense mechanisms that are based on RNA interference. The prokaryotic CRISPR-dependent antiviral activity depends on short RNA molecules (crRNAs) annealing to phage DNA or RNA fragments. Due to the maturation process of the crRNAs (Brouns et al., 2008), it is very likely that their 3' ends do not anneal with their templates, constituting a possible target for RNase T. Although appealing, this hypothesis is at present not backed up by data and therefore only speculative.

RNase T is generally regarded as restricted to γ Proteobacteria (Condon and Putzer, 2002) (although it also occurs in α Proteobacteria, e.g. *Rhodobacterales* bacterium HTCC2255, NCBI accession: ZP_01450509, revealed by blastp analysis). It was therefore of interest to see whether metagenomic environmental sequences might exist which shared common ancestry with the phage RNase T gene to better understand the origin of these genes.

4.3.8.6 RNase T: sequence comparison

The phage RNase T proteins share 47 % identity compared to 18-27 % with the top (proteobacterial) matches from NCBI (Table 4.11). Both phage genomes contain a putative bacterial RNase T, which share 47 % identity (Table 4.5). The top blastp hits of A-HIS1 and A-HIS2 RNase T were found against γ proteobacterial proteins. Sequence comparison of the *Acaryochloris* phage RNase T amino acid sequences to other DEDDh sequences from BLAST hits and other members of the DEDD exoribonuclease superfamily, showed that they are novel ((Zuo and Deutscher, 2001), Fig. A.3, Appendix A); the amino acids which give rise to the family name DEDDh are conserved, and the motifs defined by these amino acids are mostly as defined in (Zuo and Deutscher, 2001), but the inter-motif regions of the phage homologues bear little if any similarity to the other sequences.

Table 4.11: RNase T sequence comparison. Percent identity matrix comparing *Acaryochloris* phage genes with NCBI sequences using Clustal X. P = proteobacteria, R = phage RNase T. Sequence details include NCBI accession numbers.

P1	P2	P3	P4	R1	R2	%	Sequence details
100	64	62	47	27	22	P1	YP_002655118: gamma proteobacterium NOR51-B
64	100	60	46	21	24	P2	ZP_01892095: *Marinobacter algicola DG893*
62	60	100	44	23	25	P3	YP_574178: *Chromohalobacter salexigens* DSM 3043
47	46	44	100	24	18	P4	YP_002302667: *Coxiella burnetii* CbuG_Q212
27	21	23	24	100	47	R1	A-HIS1 ORF 20
22	24	25	18	47	100	R2	A-HIS2 ORF 27

4.3.8.7 RNase T: phylogenetics

To analyse the *Acaryochloris* phage RNase T sequences phylogenetically, their top BLAST sequences were acquired (all were γ proteobacterial RNase T sequences except for a *Rhodobacterales* α proteobacterial sequence and a *Haemophilus* pol III sequence) as well as their top CAMERA hits. Finally, sequences from the different DEDDh families based on the phylogenetic analyses of exoribonuclease superfamilies by (Zuo and Deutscher, 2001) were. Ambiguously aligned sequences were removed, and a DEDDy RNase D sequence from *E. coli* was used as the outgroup.

The resulting phylogenetic tree showed that the phage RNase T homologues fell alone, though were most closely related to RNase T proteins, forming a sister clade to the RNase T clade with support value 0.68 between the clades (Fig. 4.9). Interestingly, the environmental sequences fell within the RNase T clade, except for one which rooted with the outgroup. Clearly, the phage RNase T and those of the α/γ Proteobacteria share ancestry; however, the origins of the phage homologues remains unknown.

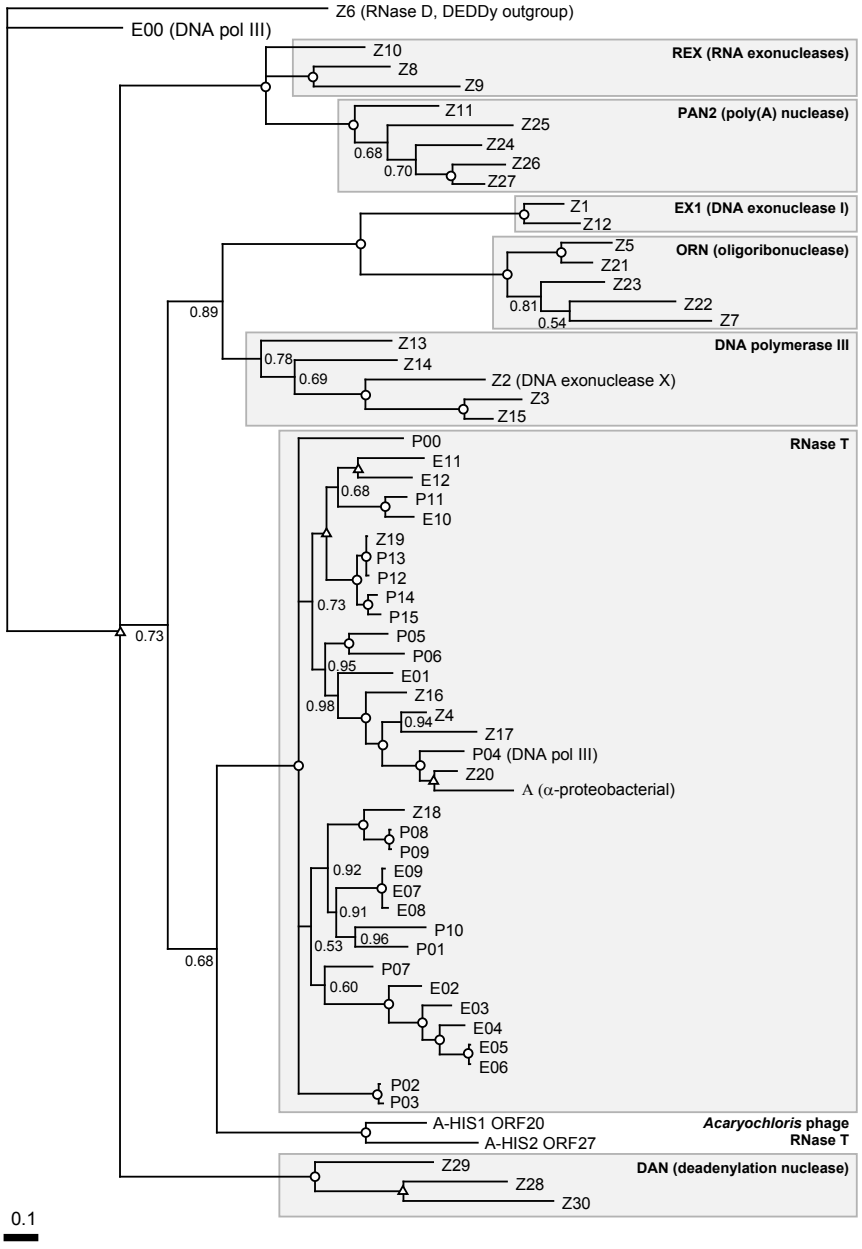


Fig. 4.9: Phylogeny of DEDD family proteins. Bayesian phylogenetic tree constructed using MrBayes. Node support values of 1.00 and 0.99 are denoted by an open circle and triangle, respectively; others are explicit. Scale bar units: amino acid substitutions per site. E=Environmental (CAMERA), P=Proteobacteria, A=Alpha proteobacteria and Z indicates they are based on the sequences that were used in the alignment by (Zuo and Deutscher, 2001). Sequence details as in Fig. A.3 (Appendix A).

4.3.8.8 Discussion: DNA polγ and RNase T

The exact origin of the eukaryotic mtDNA polymerase has remained enigmatic so far. No bacteriophage has ever been reported to infect a eukaryote and therefore it was assumed that this was not a recent gene acquisition directly from a eukaryote. Moreover, both of the genes for mtDNA polymerase and RNase T for either phage were not acquired from the cyanobacterial host strain MBIC11017, and it is unlikely that they were obtained from the original MBIC11017-like progenitor of these phages, since there was little evidence of lateral gene transfer between the strain MBIC11017 and phages A-HIS1 and A-HIS2 (§ 4.3.6.3). However, this is not surprising given that neither of these genes have ever been found in cyanobacteria. Clearly, the two *Acaryochloris* phage mtDNA polymerases branch with the mtDNA polymerase γ clade (Fig. 4.8) and this supports the hypothesis by Filée et al. that mtDNA polymerase may have originated from within a phage. However, it may be that mtDNA polymerase is the original polymerase of the protomitochondrion. It is known that many temperate phages do not possess a DNA polymerase, and as such it is possible to hypothesise the existence of a phage which contained neither mtDNA polymerase nor RNase T and that these genes were subsequently acquired by LGT.

It is widely regarded that the free-living ancestor of mitochondria is related to the α proteobacterial lineage (Gray et al., 1999), which supports the hypothesis that a mtDNA polymerase γ and RNase T were acquired from a group of bacteria related to the protomitochondrion. However, RNase T has never been found associated with mitochondria. The presence of these unique phage *rnt* (ϕ*rnt*) genes could either be a recent acquisition from undiscovered Proteobacteria-like bacteria which contain the ϕ*rnt* gene (ϕ*rnt*-bacteria), or they may have been acquired from an ancestral Proteobacterium, the progenitor of the ancestral rnt (Anc_*rnt*), which has since diverged as has rnt of extant Proteobacteria. Alternatively, both phage

genes may have been acquired either sequentially or simultaneously from an ancestral bacterium which contained both the phage mtDNA polymerase and Anc_*rnt*, and such a bacterium may have proceeded to lose either gene resulting in the hypothetical protomitochondrion-like bacteria and the extant *rnt*-containing Proteobacteria.

The ability to isolate phages on a host which is not affiliated with either the phage mtDNA polymerase or ϕ*rnt* suggests that these phages may infect (or have in the past) either protomitochondrion-like or ϕ*rnt*-bacteria. Indeed, it is unclear how these phages came to infect the cyanobacterium *A. marina*. If these hypothetical hosts are extant and susceptible to these phages this might indicate either an ancient ancestry, in which these phages infected a bacterial ancestor common to the hypothetical protomitochondrion-like bacteria, ϕ*rnt*-bacteria and *Acaryochloris*-like cyanobacteria or a recent ancestry, in which these phages may be able to infect different bacterial phyla in general. However, the latter has never been thoroughly tested or observed because in common practice phage biologists tend to limit host range experiments to the same phyla. The ancient ancestry scenario would give us some insight into the evolutionary events between the last universal bacterial ancestor and the bacterial phyla which are present today. Importantly, the ability to isolate two different phages which possess both the RNase T and the eukaryotic DNA polymerase genes indicated the occurrence of these genes in these phages was not an isolated incident. These findings have far-reaching implications concerning the co-evolution of phages, mitochondria, proteobacteria and cyanobacteria through LGT.

4.3.9 Phage replication

Genes in A-HIS1 and A-HIS2 associated with phage replication, occur in regions II and III (Fig. 4.1, Fig. 4.2). Besides VSPs, these genes include a carboxypeptidase and a deoxynucleoside kinase which are associated with

protein maturation and DNA synthesis, respectively (Alberts et al., 2002; Eriksson et al., 2002). The protease contains two conserved domains: a C39 peptidase domain associated with the processing of bacteriocins and a family 19 chitinase domain. Each phage also contains a terminase, which is involved in the packaging of DNA into the virion heads (Catalano, 2000).

4.3.10 Regulatory elements and motifs

4.3.10.1 Putative promoter of A-HIS2

Using PHIRE a putative conserved promoter sequence was detected 10 times across the A-HIS2 genome, each upstream from the start codon of a predicted ORF Fig. 4.10. The nucleotide sequence[6]:

5'-T**D**A**CC**H**N**A**K**<u>**GTTACA**</u>N**B**D**Y**<u>**GAGGTATT**</u>Y**H**Y-3'

was designated P^1 (consensus amongst the 10 putative promoter sequences in bold). Notably, there are two highly conserved regions in this promoter (underlined) which are four bases apart in all versions of the sequence. The putative promoter sequence extends a further 6 bases upstream of the P^1 region for ORFs 73, 79 and 83 at the 5' end of P^1. The nucleotide sequence 5'-**GGGGGT**P^1-3' was designated P^2. Both P^1 and P^2 were not found in any of the upstream regions of the predicted A-HIS1 ORFs.

6 out of the 10 putative promoter regions occurred upstream of putative virion structural proteins identified by mass spectrometry (ORFs 55, 57, 73, 83, 85 and 87, the tape measure protein). Interestingly, the putative promoter region also occurred upstream of ORFs 97 and 103 which code for a carboxypeptidase and putative tail fibre protein, respectively. Additionally, using ELPH one instance of a putative -10 and -35 promoter region in A-HIS1 was identified at positions 4891..4896 and 4914..4919 18 bases upstream of ORF 11, with the optimal separation of 17 bases (as identified for the major class of *E. coli* promoters (Alberts et al., 2002)).

[6]DNA base ambiguity code: D=A/G/T, H=A/C/T, K=G/T, B=C/G/T, Y=C/T, N=unknown.

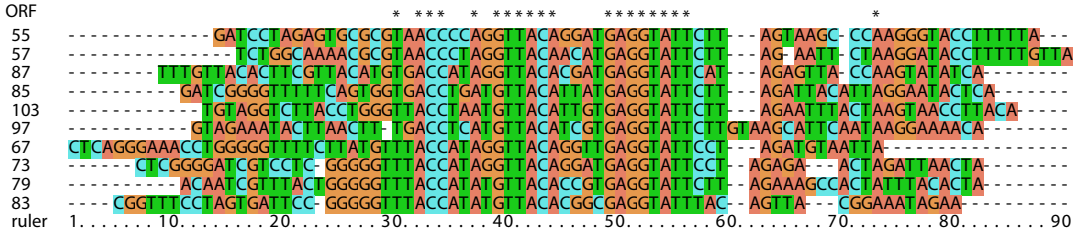


Fig. 4.10: A-HIS2 putative conserved promoter. Alignment of sequences detected in the A-HIS2 genome by PHIRE, which occurred upstream of putative predicted ORFs (ORF numbers shown on the left). The alignment was performed on the 70 bases upstream of each ORF including the first adenosine (A) at the right end of each sequence, which is the start codon of the respective downstream ORF. The alignment was performed in Clustal X using default settings. Consensus sequence are indicated by *.

The presence of the putative promoters P^1 and P^2 in A-HIS2, which occur upstream of ORFs coding for structural proteins, a carboxypeptidase and a putative tail fibre protein, indicate that these genes are probably expressed together during the replication cycle. In the phage λ replication cycle, structural proteins, like the tail fibre protein are expressed late, which suggests P^1 and P^2 are likely to be late promoters in the infection cycle (Catalano, 2005). Quantitative real-time polymerase chain reaction studies would elucidate when these genes are transcribed and when these promoters become active during the infection cycle (§ 5.3.4.4).

One-step growth experiments indicate the eclipse and lysis period is the same for both phages (Fig. 3.5) which suggests that this promoter does not affect the replication speed with regards to the first appearance of phages in the replication cycles of AHIS-1 and A-HIS2, since no similar promoter was found in A-HIS1. It is therefore unknown whether this promoter confers an advantage to A-HIS2 replication.

4.3.10.2 Programmed translational frameshifts

In dsDNA phages, it has been observed that not only is the ORF encoding the major tail protein always found upstream of the ORF which encodes the

tape measure protein, but also in between these two genes there are usually a pair of overlapping genes which possess a programmed translational frameshift (Xu et al., 2004). This has been found to be strongly conserved in tail assembly genes and the gene pair are thought to be chaperones in the process of tail assembly (Levin et al., 1993). Using BLAST and mass spectrometry, it was possible to identify the tape measure protein gene of phages A-HIS1 and A-HIS2. Examination of the upstream region of the TMP gene in each phage indicated the presence of -1 type translational frameshifts characterised by 5'-XXXYYYZ-3' 'slippery sequences'.

In the A-HIS1 genome, the slippery sequence 5'-GGGAAAG-3' was detected upstream of the initial GeneMark prediction of 47020..47193c for A-HIS1 ORF 85, and the overlap of the G/GT gene pair (where G/GT refers to the region between the genes for the major tail protein and the tape measure protein (Xu et al., 2004)) was calculated to be 48 aas, thereby extending the length of A-HIS1 ORF 85 from 57 to 104 aas (Table A.1). In the A-HIS2 genome, two slippery sequences, 5'-GGGTTTC-3' (47344..47350c) and 5'-GGGAAAG-3' (47383..47389c) were detected upstream of the initial GeneMark prediction of 47075..47251c for A-HIS2 ORF 93. The G/GT gene pair overlap for both of these slippery sequences was calculated to be 52 aas, which extended the length of A-HIS2 ORF 93 from 58 to 109 aas (Table A.2).

This data indicates that A-HIS1 ORF 86 and A-HIS2 ORF 94 are likely to be the A-HIS1/2 equivalents of the λ G gene and A-HIS1 ORF 85 and A-HIS2 ORF 93, the A-HIS2/2 equivalents of the λ T gene, where in λ, the gene order is 5'-V-G-T-H-3', where V is the major tail protein gene, G and T are the overlapping gene pair and H is the tape measure protein gene. However, unlike λ, this particular set of genes in A-HIS1/2 appear to be separated by other genes, the function of which are currently unknown. A BLAST analysis of the lengthened A-HIS1/2 T-type genes did not yield

any further information to annotate these ORFs, however, they can be preliminarily assigned the putative G/T gene function based on the current analysis. Further investigation is also required to elucidate the major tail protein.

4.3.10.3 Ribosomal binding sites

RBSfinder predicted 30 ORFs to have an RBS upstream from the original start codon for both phages using the sequence GGAGAA compared to 36 and 58 ORFs for A-HIS1 and A-HIS2, respectively, with the sequence AGGAGG derived from the two 16S genes in the *Acaryochloris* genome (see § 4.2.4 for details). There did not appear to be a significant preference of ORFs towards either consensus sequence. The RBS data is included in Table A.4 and Table A.5 (Appendix A).

4.3.10.4 Transcriptional terminators

10 and 12 transcriptional terminators were detected for A-HIS1 and A-HIS2 with transtermHP, respectively (Appendix A Table A.6). All terminators were found to be predicted on one and the same strand for both phages. Two predicted terminators for A-HIS1 were eliminated since there were ORFs in their proximity and one was a false terminator since it was intergenic. For A-HIS2 there were two pairs of overlapping predictions; (2Term8, 2Term9) and (2Term10, 2Term11). 2Term9 and 2Term8 are virtually identical in scoring and therefore it is not possible to rule out which is the most likely terminator in that position. However, 2Term11 has a higher confidence score of 93 compared to 2Term10 at 76, but only a marginally lower hairpin score, where the lower the hairpin score the more stable the hairpin (Ermolaeva et al., 2000; Kingsford et al., 2007). Therefore the true prediction of the latter pair may be 2Term11, but remains speculative. Amongst the upstream ORFs of the remaining predicted terminators were

the putative major capsid encoding ORFs 73 and 82 of phages A-HIS1 and A-HIS2, respectively.

4.3.10.5 Clustered regularly interspaced short palindromic repeats

In the *Acaryochloris* chromosome one CRISPR with three spacers (referred to here as SP1, SP2 and SP3) was found in the CRISPR database along with two questionable CRISPRs; one in each of plasmids pREB7 and pREB8, with just one spacer each - SP4 and SP5, respectively (Grissa et al., 2007b). Blasting the spacer sequences against the phage genomes gave 3 and 7 hits to phages A-HIS1 and A-HIS2, respectively (Table 4.12). In particular, SP4 and SP5 were almost identical and gave the same hits upon blasting. However, Barrangou et al. found in CRISPR studies with *Streptococcus thermophilus* that resistance is determined by the similarity between host spacer sequences and the corresponding phage sequences they were derived from (Barrangou et al., 2007). Barrangou et al. went on to conclude that a host with a particular spacer should be resistant to a phage containing an identical sequence.

Putative CRISPR		
CRISPR id(s) and length	NC_009925_1 (*A. marina* chromosome, 252 bases (b))	
SPACER 1 (SP1, 35 b)		
SP1	GGAACT-CAACGTCATGCTCTGCAATCATTTCAGTT	4815211..4815245
1S1*	AAAACTACAACGTCATGCGGCCGAGAAATTTTCGTT	18259..18294c
SPACER 2 (SP2, 35 b)		
No hits found		
SPACER 3 (SP3, 35 b)		
SP3	CTCAATTGGAAATTAGAGGAGCAAACGCAATGCAT	4815355..4815389
2S1	ATCCCAATGTTCCCGTAGGAGTAAACGCAATGCTA	52504..52538
Questionable CRISPRs		
CRISPR id(s) and length	NC_009932_1 (*A. marina*, pREB7, 136 b) and NC_009933_1 (*A. marina*, pREB8, 137 b)	
SPACER 4 (SP4, 51 b) and SPACER 5 (SP5, 52 b)		
SP4	TCTTT-GGTTTTTTCATCAAAGGCTAACGGTGCAGTTAGCCTCAATCGGTGT	59594..59644
SP5	TCTTTTAGTTTTTTCATCAAAGGCTAACGGTGCAGTTAGCCTCAATCGGTGT	90317..90368
1S2	GATAGAGATGTCAAGGTTATTTGCTAACGGTGCGACTTGTGCGATAGCGGGT	44793..44790
1S3	ACGCGAACTTGCGATCGCAAAGGCTAACGTTCCCCCCGAACTCGTCGCTCTC	36958..37009
2S2	AGCCCGAGCGGAAACCGCAAAGGCTAACCGCGATCTCTTGATCAAGACTGCG	36523..36574
2S3	TCGTACTGTGTTTTCATCAAAGGATAACCCTAGCCGTTGCATTTCACGTTGA	44830..44881
2S4	GCTTTCTCTGTTTTCATCAAAAGTCACCCTCTAGCATGTCTAGGATTTGCTC	35290..35341c
2S5	GGTTGTCTTGGCACACGCTAAGGCTAACGGGTACACCTTGGTTTCCTCCGTG	37062..37113
2S6	TGAAAGCCTAAGGAATGACGTGTTTAAC-CTGCAGTTAGCCAATGAGAACGG	42530..42580c
2S7	GGCCTCATTTGCCACCCTTACGGCTAACGGTGTAAAAACTAAGGAAGCCTCT	44932..44983c

Table 4.12: CRISPR spacer alignments. Alignment of CRISPR-spacer-associated sequences identified by standalone blastn. The same blast results were obtained for spacers 4 and 5 and so are included in the same alignment. * xSy: spacer-associated sequence y found in phage A-HISx, spacers abbreviated to SPn (n=1..5).

Small CRISPR RNAs (crRNAs) have been shown to guide antiviral defense in bacteria ((Brouns et al., 2008), § 4.3.8.5). In Chapter 3, it was observed that both phages A-HIS1 and A-HIS2 have continually growing plaques § 3.3.5. This fits with the fact the phages probably find cells in their stationary growth phase to infect. Surprisingly, after a number of weeks phage resistant colonies readily appeared in the cleared zone of plaques which suggests that *Acaryochloris* spp. may have evolved mechanisms to overcome the virulence of these lytic phages, perhaps by integrating further phage DNA fragments into their CRISPRs, since none of the phage sequences which were hits to the host spacers are identical (Table 4.12), which appears to be required for resistance (Barrangou et al., 2007). The phage resistant mutants are further examined in § 5.3.1.

4.4 Conclusion

A-HIS1 and A-HIS2 are the first phage genomes to be sequenced from siphoviruses which infect cyanobacteria, which is surprising, considering siphoviruses are the most sequenced family of the *Caudovirales*. However, given these are also the first genomes of phages which infect *A. marina* to be sequenced, it is not surprising that only 26 % (A-HIS1) and 22 % (A-HIS2) of the predicted ORFs could be assigned putative function. Indeed, the ability to assign function to predicted ORFs in A-HIS1 and A-HIS2 solely by database searches is limited by the genomic data available. This has also been observed in other phage genome studies like that of Pagaling et al. on archaeal virus BJ1 isolated from the hypersaline lake Bagaejinnor in Inner Mongolia, China (Pagaling et al., 2007). Pagaling et al. assigned function to 20 of their 70 predicted ORFs even though there were 16 archaeal viromes available at the time of their study. This study therefore highlights the importance of studying phages as well as other organisms, from different habitats.

To be sure, the *Acaryochloris* phage genomes have a unique composition and organisation when compared to other siphoviruses (Brüssow and Desiere, 2001; Seguritan et al., 2003). For instance, the phages contain a novel putative bacterial RNase T, which has never been found in a phage genome, let alone a bacterial one based on the phylogenetic tree of DEDD family exoribonucleases (Fig. 4.9), since they do not fall with environmental or current nr database sequences either. The phage RNase T genes are clearly very different from the DEDD protein sequences which are available, which is also evident from examination of the corresponding alignment (Fig. A.3, Appendix A). The phage genomes are also the first to contain mtDNA polymerase γ homologues, where a phylogenetic analysis showed that these phage sequences do not group with any other phages or bacteria with respect to this gene (Fig. 4.8). Therefore these phages cannot be easily identified with any particular *Siphoviridae* genus. Importantly, our ability to isolate two very different phages which possess these novel genes indicates the occurrence of these genes in these phages is not an isolated incident. This study therefore indicates the importance of isolating and studying niche marine phage-host systems which remain a reservoir of evolutionary history. In particular, this phage-host system uniquely provides evidence of a direct link between phages, mitochondria, Proteobacteria and Cyanobacteria.

Chapter 5

Interactions of phages A-HIS1 and A-HIS2 and *Acaryochloris marina*

5.1 Introduction

Phage-bacteria interactions often tend to be thought about in the short term, such as immediate lysis or temperate phage integration. However, long term interactions between phage and hosts are of significant relevance, for example, phages select for host resistance in cell populations (Luria and Delbrück, 1943; Mole et al., 1997; Andersson and Banfield, 2008). More directly, phages subvert host mechanisms during infection to instigate the synthesis of new progeny (Miller et al., 2003). In cyanobacteria-cyanophage systems, cyanophages have a direct influence on photosynthesis of the host to fuel their replication (Mann, 2005). During the characterisation of phages A-HIS1 and A-HIS2, apparently confluently lysed lawns of *Acaryochloris marina* strain MBIC11017 yielded spontaneous colony growth on prolonged incubation. Clonal cultures were produced from the colonies and were shown to be resistant to A-HIS1 and A-HIS2 in Chapter 3. In this chapter, these phage resistant strains are characterised using adsorption assays. Additionally, to investigate the interaction between host strain MBIC11017 and A-HIS1 and A-HIS2 further, absorbance spectroscopy and

pulse-amplitude modulation (PAM) fluorometry are used to examine the effect of infection on host photosynthesis. Finally, polymerase chain reaction (PCR) and quantitative (real time) PCR (qPCR) experiments were used to assess gene expression during infection.

5.2 Materials and methods

5.2.1 Phage resistant *Acaryochloris*

All incubations were performed at 23 °C under 10–20 μmol photons $m^{-2}s^{-1}$. Plaque assays (§ 3.2.3) which yielded confluent plates were sealed with Parafilm M™ and incubated for 1–2 months before colonies appeared. Individual sections of agar with spontaneous colonies were cut from plaque assays and placed in wells containing fresh ASW in 6-well plates. The cultures in each well were left to become turbid, and then culture from each well was streaked onto a separate agar-ASW plate and incubated. Individual new colonies were picked and streaked out onto agar-ASW plates once more before single colonies were picked and transferred to liquid media. The mutant strains which survived this process were then designated A-HISxRy, where A-HISx = A-HIS1 or A-HIS2, indicating which phage gave rise to the spontaneous strain and Ry denotes the y^{th} mutant cultured. Each stage of this process required $\sim$1 month for colonies to establish.

5.2.2 Adsorption assays

Exponentially growing *Acaryochloris* strain MBIC11017 (§ 2.2.3) (OD800 between 0.4 and 0.5 were infected with phage A-HIS1 or A-HIS2 at an MOI = 1. Samples were collected once an hour for 4 h with the first sample immediately after phage inoculation. At each time point, 0.5 mL of cells were centrifuged, transferred to a new tube and stored at 4 °C. The number of free phage in each sample was then counted by the plaque assay method

(§ 3.2.3). Unadsorbed phage were calculated as a percentage of the initial number of phage which were added.

5.2.3 Absorbance

Culture was grown and infected as in § 5.2.2. Absorbance spectra were measured as detailed in § 2.2.11. Based on the absorbance index calculation defined in (§ 2.2.11, (Six et al., 2004)), the absorbance index $I^*_{\mathrm{Chl}d}$ is defined as:

$$I^*_{\mathrm{Chl}d} = \frac{A_{716\ \mathrm{nm}}}{A*_{800\ \mathrm{nm}}} \tag{5.1}$$

where $A_{716\ \mathrm{nm}}$ is the absorbance at 716 nm at each time point and $A^*_{800\ \mathrm{nm}}$ is the absorbance at 800 nm at time point zero. The absorbance index, $I^*_{\mathrm{Chl}d}$ is subtly different from $I_{\mathrm{Chl}d}$ (§ 2.2.11, Fig. 2.3), in that it takes into account that the experiment is performed over a much shorter time period, in which the *A. marina* cells do not have sufficient time to divide and grow and therefore the number of cells over the time course is taken to be $A^*_{800\ \mathrm{nm}}$.

5.2.4 Pulse-amplitude modulation fluorometry

Culture was grown and infected as in § 5.2.2. Photosynthetic activity was periodically measured using a PHYTO-PAM Standard System I fluorometer and Phyto-Win Software version 1.45 (Heinz Walz GmbH, 2003). Samples were diluted for PAM fluorometry with fresh ASW to avoid signal overload (a total volume of 2 mL was required for the 1 cm $\times$ 1 cm quartz cuvette) and dark adapted for 3–5 min before measurements with a gain of 5 and default saturation pulse width of 0.2 s. The saturation pulse/4-channels excitation mode was used with 1 μmol photons $m^{-2}s^{-1}$ of PAR (photosynthetically active radiation). The four excitation wavelengths used are 470, 520, 645 and 665 nm. Photosynthetic activity as described by the

manufacturer can be measured as an optimal photochemical *yield* using:

$$Yield = \frac{F_m - F_t}{F_m} \tag{5.2}$$

where, F_t is the current fluorescence and F_m is the maximal value due to the saturation pulse (see (Heinz Walz GmbH, 2003) for more details). F_t is close to F_0, the minimal fluorescence yield of a dark adapted sample.

5.2.5 DNA extraction

To extract bacterial chromosomal DNA, 25–50 mL of exponentially growing cells were concentrated by centrifugation at room temperature. The cell pellet was resuspended in 0.5 mL 0.25 M Tris-HCl pH 8 with 10 mg mL^{-1} lysozyme and incubated at 37 °C for 1 h. Afterwards, 20 μL of 5 mg mL^{-1} proteinase K was added and incubated overnight at 65 °C with a final concentration of 1 % (w/v) SDS. The DNA was then extracted using the phenol/chloroform method as for phage DNA (§ 3.2.9). The DNA was resuspended in water. Phage DNA was extracted using the method detailed in § 3.2.9.

5.2.6 Polymerase chain reaction

Table 5.1 describes the contents of 25 μL PCR mixtures. Reagents were acquired from Invitrogen except for the primers (described in § 5.2.7). Reactions were performed in a Biometra TGradient or T3000 Thermocycler. The reaction cycle used is detailed in Table 5.2. At step 4 (elongation), the program reverted to step 2 (denaturation) for 35 cycles. The temperature, t, at step 3 (annealing) was varied depending on the primers using either a range of temperatures (gradient, e.g. 45–55 °C) or a single temperature. PCR products were assessed using gel electrophoresis. 1 % (w/v) agarose gels were run at 100 V for $\sim$1–2 h (see § 3.2.10 for details).

Table 5.1: Composition of 25 μL PCR mixture.

Component	Volume (μL)
Taq DNA polymerase, Recombinant	0.2
10$\times$ PCR buffer	2.5
0.2 mM dNTP mixture	2
1.5 mM $MgCl_2$	0.75
Forward primer (F)	0.1
Reverse primer (R)	0.1
DNA template	1
Water	18.35

Table 5.2: Program used for PCR reaction.

Step	Temperature (°C)	Time
1	95	20 s
2	95	20 s
3	t	20 s
4	70	30 s
5	70	2 min
6	4	Pause

5.2.7 Polymerase chain reaction: primer design

PCR primers were designed in Primer3 (v.0.4.0) and obtained from VH Bio Ltd. (Rozen and Skaletsky, 2000). PCR primers used in this study are listed in Table 5.3 and were diluted with water to 0.1 mM before use.

5.2.8 RNA extraction

Surfaces were cleaned with RNase AWAY ® (Molecular BioProducts). 50 or 100 mL of exponentially grown cell culture was used for RNA extraction. Samples were collected, centrifuged, resuspended in 0.75 mL TRIzol ® (Invitrogen) and frozen with liquid nitrogen. After all samples were collected they were defrosted for RNA extraction (method based on that described in (Clokie et al., 2006a)). Samples were resuspended with another 0.75 mL TRIzol ® and repeatedly pipetted to make the mixture as homogeneous as possible. All samples were then heated in a water bath for 20–40 min at 65 °C, 0.3 mL chloroform was added and the samples were shaken for 30 s, left at room temperature for 10 min and then centrifuged at 15,000 g for

Table 5.3: PCR primers.. F = forward, R = reverse, 1 = phage A-HIS1, 2 = phage A-HIS2, F/R*x*, *x* = ORF number (Table 4.2 and Table 4.3).

Gene	Primer name	Primer sequence (5'-3')	Product size (b)
Host			
RNA polymerase β subunit	rpoBF	TGG GTA AAG TTA CGC CCA AG	245
	rpoBR	CTT GTC ACC CAC CTG GAG TT	
Phage			
Putative major capsid, ORF 73 (A-HIS1) and ORF 82 (A-HIS2)	mcapF	GAA SGC ACT ACC CGW CAA TC	258
	mcapR	ATC RAC GCC AGC RTA MGT AT	
mtDNA polymerase γ, A-HIS1 ORF 14	1F14	CGA ACG TCG AGG TAA AGA GC	237
	1R14	AAG GTT CCG GTG AGT CCT TT	
RNase T, A-HIS1 ORF 20	1F20	CGG CGA GAA TTC CGT ATA AA	151
	1R20	CCG GAT TTG ATA AGC GAG AG	
Putative tail fibre protein, A-HIS1 ORF 95	1F95	GTT AAC CTC GCC GTT GAC AT	223
	1R95	CGT CCG CTA GAC CAG AAA AG	

10 min. The aqueous layer ($\sim$1 mL) was removed to a new Eppendorf tube to which 0.5 mL of isopropanol (4 °C) was added and gently inverted. This mixture was incubated for 12 min on ice and centrifuged at 15,000 g for 10 min at 4 °C to pellet the RNA. The supernatant was removed and the pellet was washed with 0.5 mL 75 % (v/v) ethanol (diluted from 100 % with RNase-free water (Sigma)) and centrifuged once more at 4 °C. The ethanol was removed with a pipette and left to air dry for $\sim$15 min.

5.2.9 RNA purification

RNA pellets were prepared as in § 5.2.8 and resuspended in 90 μL RNase-free water and 4 μL of Ambion TURBO^{TM} DNase (2 U mL^{-1}) with 10 μL 10$\times$ DNase buffer. Samples were then left from 20–30 min at 37 °C. A Qiagen RNeasy Mini Kit was used to purify the RNA sample as described by the manufacturer. For the final step, the RNA sample was eluted from the column with 50 μL of RNase-free water heated to 50 °C in a water bath. To check for DNA contamination, PCR was used as described in § 5.2.6. If a PCR product could be amplified using phage and host specific primers, the RNA purification procedure was repeated until no such PCR product could be detected.

5.2.10 Complementary DNA synthesis

400–1000 ng of RNA was used per 20 μL cDNA (complementary DNA) reaction. Each reaction was made up of the components described in Table 5.4. The RNA, RNase-free water (Qiagen), dNTPs and random hexamers were combined for each reaction using master mixes and heated at 70 °C for 5 min followed by 10 min at 4 °C using a Biometra TGradient or T3000 Thermocycler. The remaining components were made into a master mix and added to each reaction and the samples were heated at 25 °C for 5 min, 50 °C for 60 min and 70 °C for 15 min. Samples were then stored at -80 °C.

Table 5.4: Composition of cDNA reaction. +VH Bio Ltd.,* Invitrogen.

Component	Volume (μL)
RNA	x
RNase-free water	12-x
10 mM dNTP mix*	1
$\sim$150 μM random hexamers+	1
5$\times$ First Strand Buffer*	4
0.1 M DTT*	1
SuperScript^{TM} III Reverse Transcriptase (RT)*	1

5.2.11 Quantitative polymerase chain reaction

5.2.11.1 Primer design

Quantitative (real-time) polymerase chain reaction (qPCR) primers were designed using ABI PRISM^{TM} Primer Express^{TM} version 2.0.0 and the TaqMan ® Probe & Primer Design option (Applied Biosystems^{TM}). Default parameter requirements; primer melting temperature, T_m: minimum T_m = 58 °C, maximum T_m = 60 °C and optimal T_m = 59 °C; primer GC content: minimum GC % = 30, maximum GC % = 80; primer length: minimum length = 9 bases, maximum length = 40 bases, optimal length = 20 bases; amplicon: minimum T_m = 0 °C, maximum T_m = 85 °C, minimum length = 50 bases, maximum length = 150 bases. qPCR primers were acquired from Invitrogen^{TM} (Table 5.5).

5.2.11.2 Sample collection and plate setup

A. marina was grown to the exponential phase (using the same conditions as § 2.2.3) in 4 L cultures in glass jars under sterile conditions. These cultures were supplemented with 6.7 mL of 7.5 % (w/v) $NaHCO_3$ per litre of ASW before their growth was monitored. For each experimental replicate, 1 L was infected with A-HIS1, 1 L was infected with A-HIS2 and 1 L was uninfected in sterile 1 L conical flasks. The phages were inoculated to an MOI

Table 5.5: qPCR primers.. F = forward, R = reverse. The number following F or R indicates the phage by the first digit (1:A-HIS1, 2:A-HIS2) and then the ORF number by the subsequent digits (see Table 4.2 and Table 4.3) with the exception of *rpoB* which refers to the host *A. marina* RNA polymerase B gene.

Gene	Primer name	Primer sequence (5'-3')	Product size (b)
Host *A. marina*			
RNA polymerase β subunit	FrpoB	CTC CGG CAA CTG GAT GAG A	62
	RrpoB	ATG TCG CCG GCT TCG A	
Phage A-HIS1			
mtDNA polymerase γ	F114	CGA AAA GCG GAT CAA CCT TCT	68
	R114	TGC GGG TAG CGA GTC CAA	
RNase T	F120	GCG ATG GAA ATC AAC GGT TT	68
	R120	CGC CAG ATT TGC GAA AGC	
DNA pol I-like FEN	F130	CGG GTT ACC CTC CGT TAC TCA	68
	R130	TTG CGC GGC GTT AAA GA	
Putative major capsid	F173	CGG TAT GCT GTT ACA GGA CGA A	67
	R173	AGC GGC GGC GAT AGA GT	
Putative tail fibre protein	F195	CGG TGT GAG ATC GCC TTT TT	65
	R195	TGA CGA ATC GGG CGA AA	
Phage A-HIS2			
mtDNA polymerase γ	F220	CAA GGG TTC CTC GTA CTG CTA TG	65
	R220	GAC GAC TGC GGG CCT TAA	
RNase T	F227	CTG TGA TTT CCA CGG GAT CAC	73
	R227	TAC GGT ACG CTC TGC ATC GT	
Putative major capsid	F282	CGT TCA GTC GGG ACC AAT G	67
	R282	CCG CAG GAC CAG CGA TT	
Putative tail fibre protein	F2103	GGG ATG TAC CTT GAT GAC ACC AA	79
	R2103	TCT CAG TAG TAT TAG GCC CGA GAA A	

of 1. Cultures were shaken at 70 rpm under 30 μmol photons $m^{-2}s^{-1}$. 100 mL of sample was collected per infected culture 10 minutes prior to 2, 4, 6 and 8 h after infection. These samples were centrifuged in a Hettich Rotina 46R centrifuge at 4,754 g for 15 min at 4 °C. The supernatant was poured away and the samples were treated as in § 5.2.8 to extract RNA, which was then purified (§ 5.2.9). RNA was quantified using a NanoDrop® ND-1000 spectrophotometer. cDNA was prepared as in § 5.2.10. qPCR reactions were prepared as in Table 5.6 in 96 well plates by combining a SYBR ® Green PCR Master Mix, nuclease free water (Qiagen) and qPCR primers to 150 nM (concentration of primers in Table 5.5 were optimised by assessing standard curves § 5.2.11.3). Finally template was added, i.e. cDNA (or genomic DNA for standard curves § 5.2.11.3). No template controls were carried out for each primer set.

Table 5.6: Composition of qPCR reaction. *The primer stock concentration was 3.75 pmol μL^{-1}.

Component	Volume (μL)
SYBR ® Green PCR Master Mix	12.5
Forward primer*	1
Reverse primer*	1
Nuclease free water	9.5
Template	1

Plates were subsequently sealed with adhesive covers (ABgene ®) and analysed using an Applied Biosystems™ 7500 Fast Real-Time PCR System. The program used was: 50 °C for 2 min, 95 °C for 10 min and 40 cycles of 95 °C for 15 s. This was followed by 95 °C for 15 s, 60 °C for 20 s, 95 °C for 15 s and 60 °C for 15 s at the end of each run to perform a melting curve analysis (melting curves for different genes included in Appendix A, Fig. A.4). Results were analysed using Applied Biosystems™ Sequence Detection Software version 1.4 (7500 Fast System). Intra-plate variation was monitored by including triplicate reactions for each of A-HIS1 ORF 14, A-HIS1 ORF 20 and A-HIS1 ORF 30 with a fixed amount of A-HIS1 genomic

DNA (5 ng).

5.2.11.3 Primer optimisation

For primer optimisation, standard curves were used to check that the amplification was even across different concentrations of DNA at a primer concentration of 150 nM for all primers. A serial dilution was performed with a known mass of genomic DNA (phage or host), m_{DNA}. The dilutions ranged between 150 ng at the most concentrated and $\sim$70 fg for host genomic DNA and $\sim$0.5 fg for phage genomic DNA (lowest concentration corresponds to $\sim$ 8 gene copies, § 5.2.11.5). Standard curves (scatter plots) of C_T values vs. log (concentration of genomic DNA)) were generated with either genomic DNA from the host *A. marina* for host genes or phage genomic DNA from A-HIS1 or A-HIS2 for A-HIS1 or A-HIS2 genes, respectively. C_T is the cycle value at a certain threshold in the exponential phase of the fluorescence signal from the SYBR ® Green in the qPCR reaction. Standard curve equations and corresponding R^2 values for each gene were calculated in Microsoft ® Office Excel. Amplification efficiency, E (%), was calculated using

$$E = (10^{-1/m} - 1) * 100 \tag{5.3}$$

where m is the slope value obtained from the line of best fit of the standard curve. R^2, the coefficient of determination, is a measure of how close the line of best fit corresponds to the actual data. A slope value between -3.0 and -3.9 is generally accepted and R^2 should be greater than 0.985 (Sigma-Aldrich Co., 2008). Similarly, the amplification efficiency, E, should be between 80–110 %.

5.2.11.4 Absolute quantitation

The standard curves were also used for absolute quantitation. The equation of the standard curve takes the form:

$$C_T = m * x + c_i \tag{5.4}$$

where C_T is the threshold cycle value for a particular gene and m is the slope of the line of best fit (see § 5.2.11.3). x is log (DNA concentration (ng)) and c_i is the y-axis intercept. The intercept c_i takes a different value if x is converted from log(DNA concentration) to log(gene copy number) (§ 5.2.11.5), but all other parameters of the standard curve remain the same. Therefore, given a C_T value, a copy number can be computed. The parameters for the different forms of the standard curve for each gene tested are detailed in Table 5.7.

Table 5.7: Standard curve parameters. *m* is the slope of the line of best fit of the standard curve, R^2 is the coefficient of determination. The y-intercept, c_i is c_1 for *x* = log (DNA concentration) and c_2 for *x* = log (gene copy number) (see Equation 5.4, § 5.2.11.3). E is the amplification efficiency (Equation 5.3).

Gene	*m*	c_1	c_2	R^2	E (%)
Host *A. marina*					
rpoB	-3.3833	20.2793	33.3242	0.9929	97.50
Phage A-HIS1					
ORF 14	-3.4816	12.6157	37.7344	0.9987	93.74
ORF 20	-3.5115	11.6785	37.0132	0.9953	92.65
ORF 30	-3.1472	10.6148	33.3207	0.9992	107.85
ORF 73	-3.8704	9.9160	37.8402	0.9915	81.29
ORF 95	-3.6871	10.2111	36.8123	0.9929	86.73
Phage A-HIS2					
ORF 20	-3.1561	13.2216	35.9495	0.9993	107.42
ORF 27	-3.4237	12.0651	36.7205	0.9987	95.92
ORF 82	-3.1191	12.5364	34.9981	0.9998	109.22
ORF 103	-3.2899	12.4476	36.1390	0.9996	101.35

5.2.11.5 Mass of DNA to gene copy number

To convert a mass of DNA (based on (Biosystems, 2003)), m_{DNA}, to gene copy number, N_c, the mass of a genome was first calculated for the host *A. marina* (8,361,599 bp) and each phage (A-HIS1: 55,653 bp, A-HIS2: 57,391 bp). The mass of a genome, m_{genome} (g), was calculated using

$$m_{genome} = \frac{\gamma * M_{DNA}}{N_A} \tag{5.5}$$

where γ (bp) is the genome size, M_{DNA} is the average molecular weight of a double stranded DNA molecule = 660 g mol^{-1} and N_A is Avogadro's constant (6.022×10^{23} mol^{-1}). The genes of interest are single copy genes in the genome in which they are found, therefore

$$N_c = \frac{m_{DNA}}{m_{genome}}. \tag{5.6}$$

5.3 Results and discussion

5.3.1 Phage resistant *Acaryochloris*: adsorption

Plates which showed confluently lysed lawns using strain MBIC11017 infected with either phage A-HIS1 or A-HIS2 yielded dense colony growth after 1–2 months of being left in the incubator (Fig. 5.1). Single colonies were transferred to liquid medium. However, none of these grew. Therefore, colonies were subsequently made clonal by growing them on plates over approximately 6 months using the method detailed (§ 5.2.1). This work produced 5 strains designated A-HIS1R1, A-HIS1R2, A-HIS1R3, A-HIS2R1 and A-HIS2R2 (Table 3.2).

In Chapter 3, it was noted that strain CCMEE5410, and the resistant phage strains did not lyse in the presence of A-HIS1 and A-HIS2 in well assays (§ 3.3.7). Adsorption assays were used to assess whether these phages

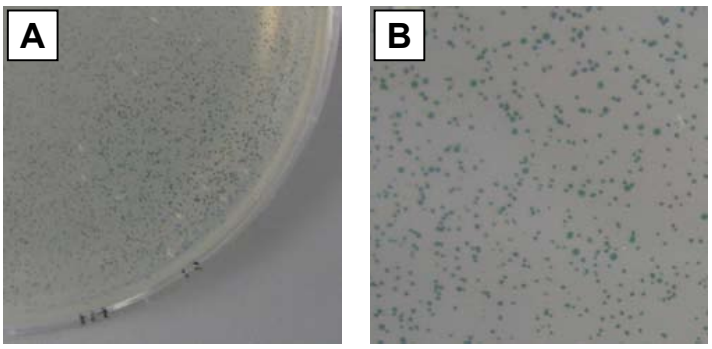


Fig. 5.1: Phage resistant *Acaryochloris*. (A) Example of a confluently lysed lawn from infection of strain MBIC11017 with phage A-HIS1, which yielded spontaneous colony growth and (B) close up of colonies.

were able to attach to selected hosts: the free-living strain CCMEE5410 (Miller et al., 2005) and phage resistant strains A-HIS1R1 and A-HIS2R1. To monitor the health of strains MBIC11017, CCMEE5410, A-HIS1R1 and A-HIS2R1 during infection, the OD800 (optical density at 800 nm) of infected cultures (with either A-HIS1 or A-HIS2) and an uninfected control containing no phage was measured. The adsorption experiment was also performed on strain MBIC11017 for comparison, which was the only strain that the two phages formed plaques on in plaque assays. Adsorption experiments were performed with MOI=1 (multiplicity of infection or ratio of phage to bacteria). In Fig. 5.2, unadsorbed phage were calculated as a percentage of the initial number of phage which were added, which is represented as the zero time point at 100 % in each graph based on stock titre. In each graph, the second time point is at 5 min.

Clearly, both phages A-HIS1 and A-HIS2 adsorbed to the host MBIC11017 after 5 min with only $\sim$10 % and $\sim$1 % of unadsorbed phage remaining in the supernatant, respectively (Fig. 5.2(A)). After infection, the OD800 changed significantly in the case of strain MBIC11017, where it decreased in infected cultures compared to the control (Fig. 5.3(A)). A distinct increase in the percentage of unadsorbed phage for A-HIS2 was observed between 1 h and 2 h and onwards compared to 2 h and 3 h and onwards for phage A-

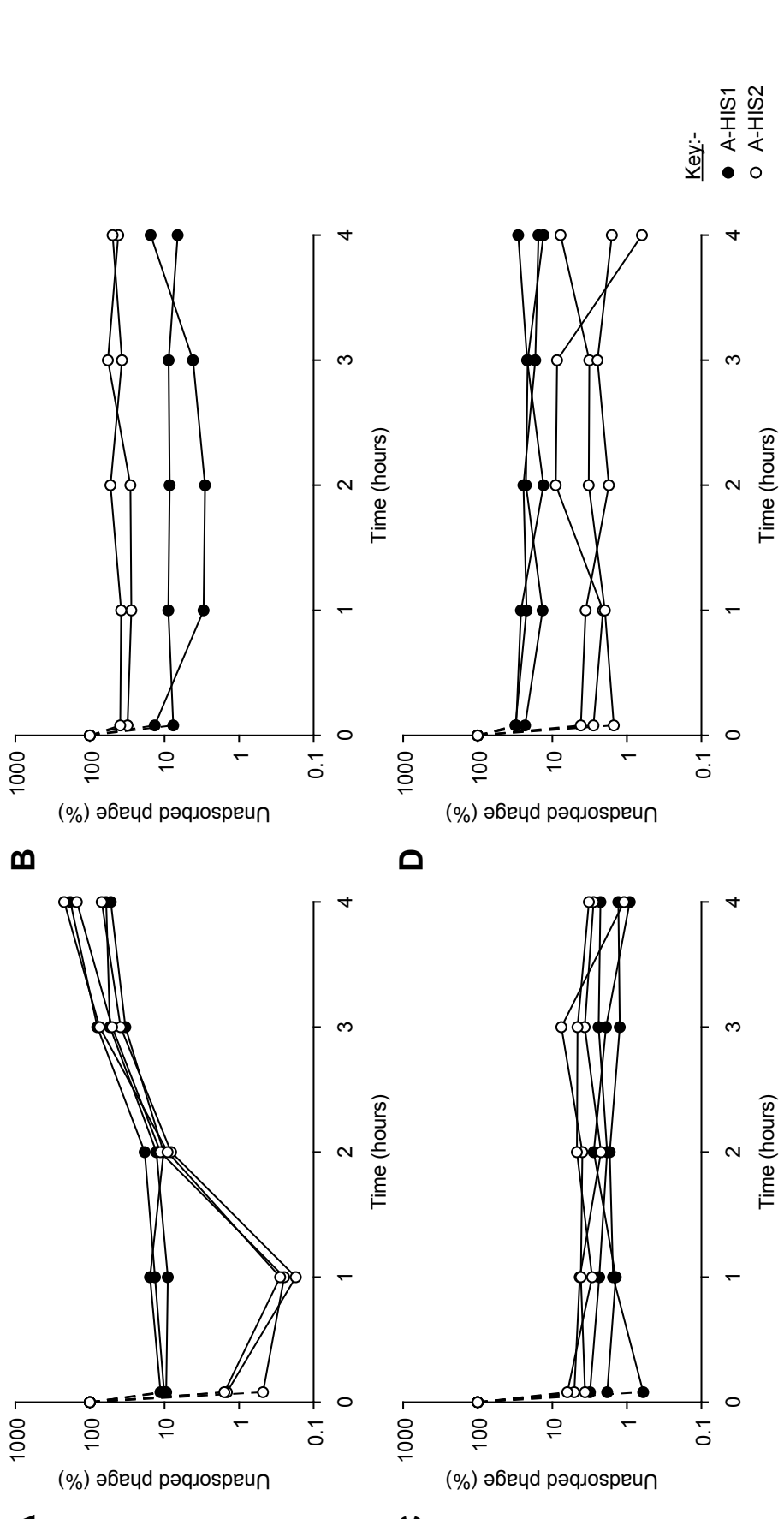


Fig. 5.2: Phage adsorption to *Acaryochloris* strains. Percentage of free unadsorbed phage as determined by plaque assay. *Acaryochloris* strains (A) MBIC11017 (wild-type parent strain), (B) CCMEE5410 (non-host), (C) A-HIS1R1 (spontaneous strain from A-HIS1 infection) and (D) A-HIS2R1 (spontaneous strain from A-HIS2 infection) were inoculated at MOI=1 (experimental replicates shown in each graph - A/C/D: n=3; B: n=2). Dotted lines join zero time-point (known titre value before experiment) to the first time-point at 5 min. Subsequent experimental time points are joined with solid lines.

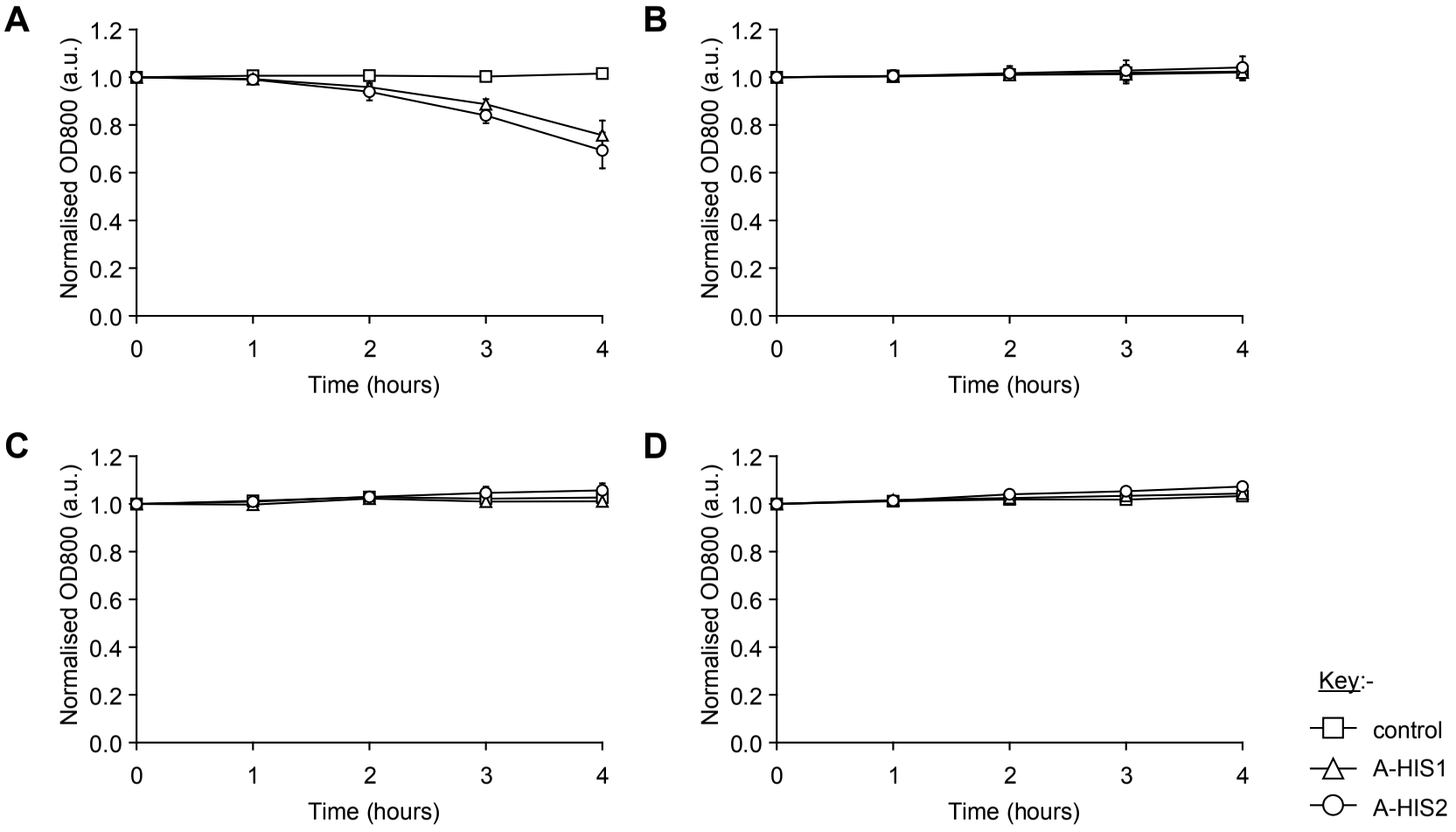


Fig. 5.3: Phage adsorption to *Acaryochloris* strains. Data were normalised at time point zero (OD800 is optical density at 800 nm) for strains (A) MBIC11017, (B) CCMEE5410, (C) A-HIS1R1 and (D) A-HIS2R1 upon infection with A-HIS1, A-HIS2 or neither (control). MOI=1, error bars are one standard deviation (n=2).

HIS1. Notably, there is a stark contrast between the adsorption experiment with strain MBIC11017 and the one-step growth curves of the two phages (Fig. 3.5). Namely, that an increase in phage can be seen between 1 h and 2 h for phage A-HIS2 compared to 2 h and 3 h for A-HIS1, suggesting that phage may have already replicated within 1 h and 2 h for phages A-HIS2 and A-HIS1 respectively. These differences are likely to be caused by variations in the amounts of phage used in the adsorption and the one-step growth experiments, which had an MOIs of 1 and 0.1, respectively. The one-step growth experiments also included a synchronisation of phage infection, whereas the adsorption experiments did not.

The same adsorption experiment on strains CCMEE5410, A-HIS1R1 and A-HIS2R1 showed that phages A-HIS1 and A-HIS2 attach to all three strains to varying degrees (Fig. 5.2(B/C/D)). However, the OD800 was observed not to change significantly for any of these host/phage combinations suggesting the cultures remained healthy after phage inoculation (Fig. 5.3(B/C/D)). A-HIS1 was shown to have a greater affinity for binding to strain CCMEE5410 than A-HIS2, where up to 57 % of phage A-HIS2 were present as free phage, compared to the maximum of 15 % of unadsorbed A-HIS1.

Interestingly, the percentage of unadsorbed phage on strain CCMEE5410 and the spontaneous strains A-HIS1R1 and A-HIS2R1 were consistent and remained at a constant level for each phage compared to the wild-type MBIC11017 (Fig. 5.2(C/D)). For A-HIS1R1, a strain which originated from a confluent A-HIS1-infected *A. marina* lawn, the percentage of unadsorbed phage was similar for both phages and ranged from 1–3 % for A-HIS1 and 1–8 % for A-HIS2 (Fig. 5.2(C)). However, the means of the percentage of unadsorbed phage were shown to be significantly different between A-HIS1 and A-HIS2 by use of the Student's t-test on the 5 min time point (t-value = 2.99 $>$ critical t-value = 2.77; same method as calculations in Table 5.8).

For strain A-HIS2R1, which originated from a confluent A-HIS2 infection, 13–31 % of A-HIS1 were retrieveable compared to 1–9 % of A-HIS2.

To determine whether a particular phage binds equally to a resistant strain compared to the wild-type strain MBIC11017, the Student's t-test was used to compare the 5 min time point from the different adsorption experiments (Table 5.8). The t-statistic was greater than the critical t-value in all cases except for the case of phage A-HIS2 and strains MBIC11017 and A-HIS2R2, where the critical value was greater than the t-statistic, and so the null hypothesis that the samples had equal means was accepted, i.e. phage A-HIS2 binds equally to strains MBIC11017 and A-HIS2R2. However, A-HIS1 which was used to isolate the resistant strain A-HIS1R1 binds better to it than the wild-type. These results indicate A-HIS1R1 and A-HIS2R1 are essentially different in their binding to phages A-HIS1 and A-HIS2 compared to the wild-type and to each other. In particular, the phage against which the resistant strains were not isolated binds less well than to the wild-type MBIC11017. It should be noted that A-HIS1 also appears to bind equally to the wild-type strain MBIC11017 and the free-living strain CCMEE5410, while A-HIS2 binds considerably less well to strain CCMEE5410 than to MBIC11017. These results imply the difference in adsorption between the wild-type MBIC11017 and the other strains is due to a difference between the host receptors to which the phage bind.

Table 5.8: Phage adsorption Student's t-test statistics. Calculation of Student's t-statistic and associated p-value (probability of getting the observed t-value by chance) based on the percentage of unadsorbed phage; two-sample assuming equal variances comparing 5 min time point values of strain MBIC11017 with that of A-HIS1R1 (n=3) and A-HIS2R1 (n=3) for a particular phage. The critical t-value is 2.78 (two-tail) and significance level, α = 0.05. The means of two data sets are significantly different if the calculated t value exceeds the critical t-value. Computed in Microsoft ® Excel.

Strain	A-HIS1R1		A-HIS2R2	
Phage	A-HIS1	A-HIS2	A-HIS1	A-HIS2
t	9.38	4.57	6.83	1.96
p	0.0007	0.01025	0.0024	0.1210

5.3.2 Phage resistant *Acaryochloris*: discussion

The occurrence of spontaneous phage resistant colonies after prolonged incubation of a host with phage has been previously observed (Luria and Delbrück, 1943; Mole et al., 1997). Luria and Delbrück did an extensive study on phage resistance using *E. coli* and phage T1 and noted that the phages which gave rise to the resistant strains generally were unable to adsorb to the new strains compared to the phage-sensitive bacteria used to produce them. In their study, Luria and Delbrück tested two main hypotheses: (1) the mutation hypothesis, that phage resistance was due to spontaneous mutation in bacteria before exposure to phage and (2) the acquired hereditary immunity hypothesis, that resistance was acquired during the attack of the virus. The results of their experiments indicated that the number of resistant bacteria in similar cultures varied and that the distribution was not random. Therefore, they concluded that phage resistance was independent of the exposure to phage and was instead due to spontaneous mutations which were present before infection with virus (Luria and Delbrück, 1943).

Much more recently and in contrast to the above results, Mole et al. showed in the host-phage system consisting of the freshwater cyanobacterium *Anabaena* sp. strain PCC7120 and the cyanophage AN-15, two types of spontaneously resistant mutants arose: the first type were ones in which the cyanophage could no longer attach to the cells and the second type involved those to which attachment could occur but the interaction resulted in defective replication (Mole et al., 1997). Using Southern hybridisation experiments Mole et al. did not detect AN-15 DNA in their resistant strains and concluded that the mutants were not likely to be lysogens and therefore, were likely to have arisen from spontaneous mutation. In light of the observations from the *Acaryochloris* phage genomes (Chapter 4), the resistant strains are unlikely to have been converted to lysogens by A-HIS1 or

A-HIS2 since the A-HIS1 and A-HIS2 genomes do not contain a lysogeny module (Chapter 4). Evidently, the host *A. marina* genome contains numerous integrase genes as well as two proteins with a λ cI motif (Swingley et al., 2008). However, nothing is currently known about these genes in terms of their activity or whether they are non-functional phage relics.

In contrast to Luria and Delbrück and in part to Mole et al., the *Acarychloris* phages which gave rise to the resistant strains are able to adsorb to the new strains. Clearly, both phages are also able to attach to the wild-type strain MBIC11017 and also the free-living strain CCMEE5410. In Chapter 4, the study of the A-HIS1 and A-HIS2 genomes showed that the phages have many genes in common between them (§ 4.3.6.1). Presumably, one of these encodes the phage tail receptor protein which is responsible for phage adsorption in A-HIS1 and A-HIS2 (Fig. 3.3(G/H/I/J)). However, while the genes between the phages may be similar, no genes are identical (Table 4.5 and Table 4.6), which may account for the differences observed in adsorption between the two phages to the different strains. This may also go some way to account for the fact that resistance to A-HIS1 conferred resistance to A-HIS2 and vice versa in all resistant strains isolated.

While it is intriguing that the resistant strains have different binding affinities to the phage compared to the wild-type based on the adsorption experiments, there is no evidence as yet to indicate the origin of the putative receptor mutations. The putative receptor mutations may have occurred by spontaneous mutation of the wild-type host receptors and were present in the wild-type culture before infection or they may have appeared after interaction with the *Acaryochloris* phages *sensu* the mutation/acquired hereditary immunity hypotheses of Luria and Delbrück. Indeed, it is not known how these differences in host receptors might confer resistance to the phages, since the phages are still able to readily attach to the resistant strains (although to varying degrees); perhaps the injection mecha-

nism of viral DNA into the host is affected. In light of these observations and those of the genomes of phages A-HIS1 and A-HIS2 in Chapter 4, the resistance could be due to phage restriction (Wood, 1966) or perhaps, these resistant strains may have integrated further phage DNA fragments into their CRISPRs (§ 4.3.10.5) to overcome the virulence of phages A-HIS1 and A-HIS2.

5.3.3 Phage-induced host modification

5.3.3.1 Absorbance

During the creation of stock lysates, it was observed that strain MBIC11017 exhibited certain physiological changes after infection with phages A-HIS1 and A-HIS2. To investigate the observations, exponentially growing strain MBIC11017 culture was inoculated at a MOI of 1, and the culture pigmentation was observed to change dramatically. In particular, cells were observed to visibly clump in shaken liquid culture typically 3–4 h after infection. In a natural biofilm, increased cell-cell adhesion would presumably make it more likely for an infected cell to remain there, thereby increasing the chances of phages spreading to other cells in the biofilm, however, this remains speculative.

Changes in pigmentation were clearly observed from absorbance spectra (Fig. 5.4), which showed significant reduction in the Chl *d* (716 nm), phycobilin (640 nm) and carotenoid (500 nm) characteristics as well as an overall decrease in absorbance over time. Analysis of the absorbance index, $I^*_{\mathrm{Chl}d}$, which gives a measure of Chl *d* per cell over the infection cycle, showed that the index decreased over the infection cycle for both phages, with the index decreasing slightly quicker on average for A-HIS2 than A-HIS1, which suggests that MBIC11017 cells infected with A-HIS2 may not be as photosynthetically active as those infected with A-HIS1. This is further investigated with PAM fluorometry and discussed in § 5.3.3.2.

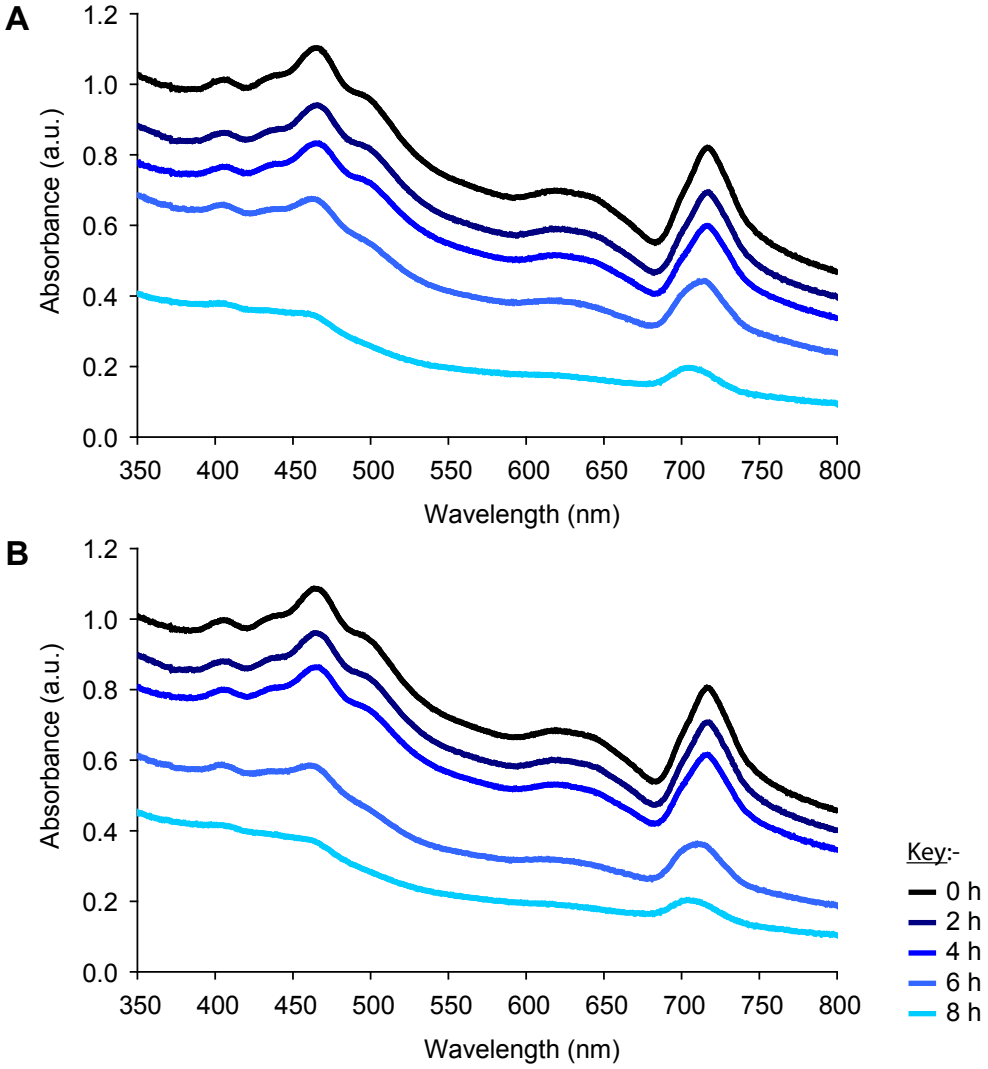


Fig. 5.4: Effect of phage on absorbance. Phages (A) A-HIS1 and (B) A-HIS2 inoculated on strain *A. marina* MBIC11017 at a MOI=1.

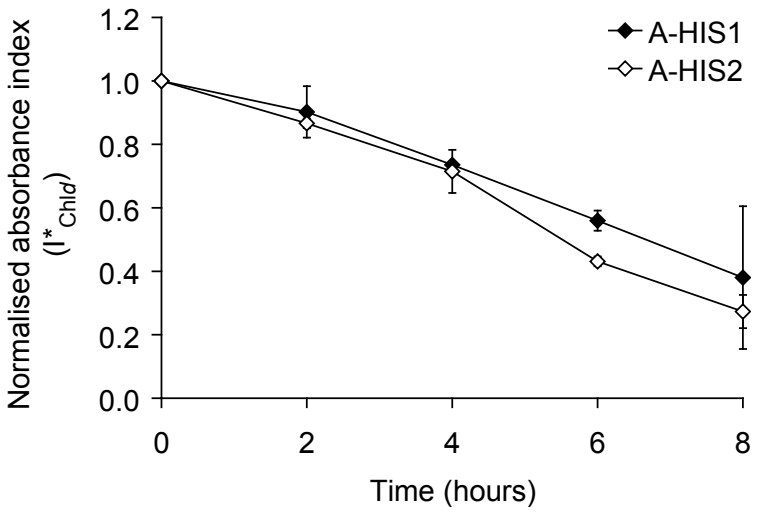


Fig. 5.5: Effect of phage infection on absorbance index. Phages A-HIS1 and A-HIS2 inoculated on strain *A. marina* MBIC11017 at a MOI=1. The absorbance index I^*_{Chld} = $A_{716\,nm}/A^*_{800\,nm}$, where $A_{716\,nm}$ is the absorbance at 716 nm at each time point and $A^*_{800\,nm}$ is the absorbance at 800 nm at time point zero. The absorbance index was then normalised to the zero time point (n=2). See § 5.2.3 for more details.

It has previously been shown that cyanophages can influence the physiology of the host during infection. For example, the well-characterised myovirus S-PM2 has been shown to induce an increase in the synthesis of the phycobiliprotein phycoerythrin in *Synechococcus* sp. WH7803 (Shan et al., 2008). The changes seen in the absorbance spectra of strain MBIC11017 during infection suggest the host may be modified in two ways. Firstly, a reduction in the characteristics of the whole cell spectra during infection which correspond to different pigments suggests that the pigments are likely to be broken down over the infection cycle and contribute to phage protein synthesis in contrast to S-PM2 infection of *Synechococcus* sp. WH7803, where data indicated that the phycobilisomes were lengthened (Shan et al., 2008). Secondly, the overall reduction of the entire spectra over time suggests the cells may absorb less light. However, considering the adsorption assays performed on strain MBIC11017 in Fig. 5.2A, lysis for phages A-HIS1 and A-HIS2 at the higher MOI of 1 may occur as soon as 1–2 h and 2–3 h after infection, respectively, which may significantly

contribute to the reduction in the (infected) whole cell absorbance spectra as cells are burst open.

5.3.3.2 Pulse-amplitude modulation fluorometry

To investigate the effect of phage on the photosynthetic capacity of the host strain MBIC11017, PAM fluorometry was used. Exponentially growing culture was inoculated at an MOI of 1. This showed a significant decrease in the quantum yield (of energy conversion in photosystem II) after phage infection with either A-HIS1 or A-HIS2 (Fig. 5.6(A/B)). Typically, samples have a yield between 0.5–0.8 ((Heinz Walz GmbH, 2003)). It was found that using the same culturing conditions (§ 2.2.3), the starting (pre-infection) quantum yield of a culture grown to exponential phase varied considerably between experiments (between 0.3 and 0.7). Non-infected culture was also monitored (Fig. 5.6(C)), for which the quantum yield remained largely constant over the time course, no matter what the starting quantum yield. All data from the different treatments were normalised and combined to compare them, which showed that while the quantum yield of non-infected culture did not change much, the photosynthetic capacity of strain MBIC11017 was reduced quicker with infection by A-HIS2 than A-HIS1 (Fig. 5.6(D)).

Previously, yield values of 0.4–0.6 have been reported as typical of wild-type *Synechococcus* cells (Campbell et al., 1998), though levels lower than this have been observed, for example in the case of *Synechococcus* WH7803, where the maximal photochemical efficiency was less than 0.3 (Clokie et al., 2006a). A similar study was also performed on the eukaryotic phytoplankton *Micromonas pusilla* (Brussaard, 2004), where the maximum quantum efficiency was $\sim$0.4 for non-infected algal cultures and a steady drop of photochemical efficiency was observed in infected cultures. In considering the apparent variation in starting quantum yield, it should be noted that

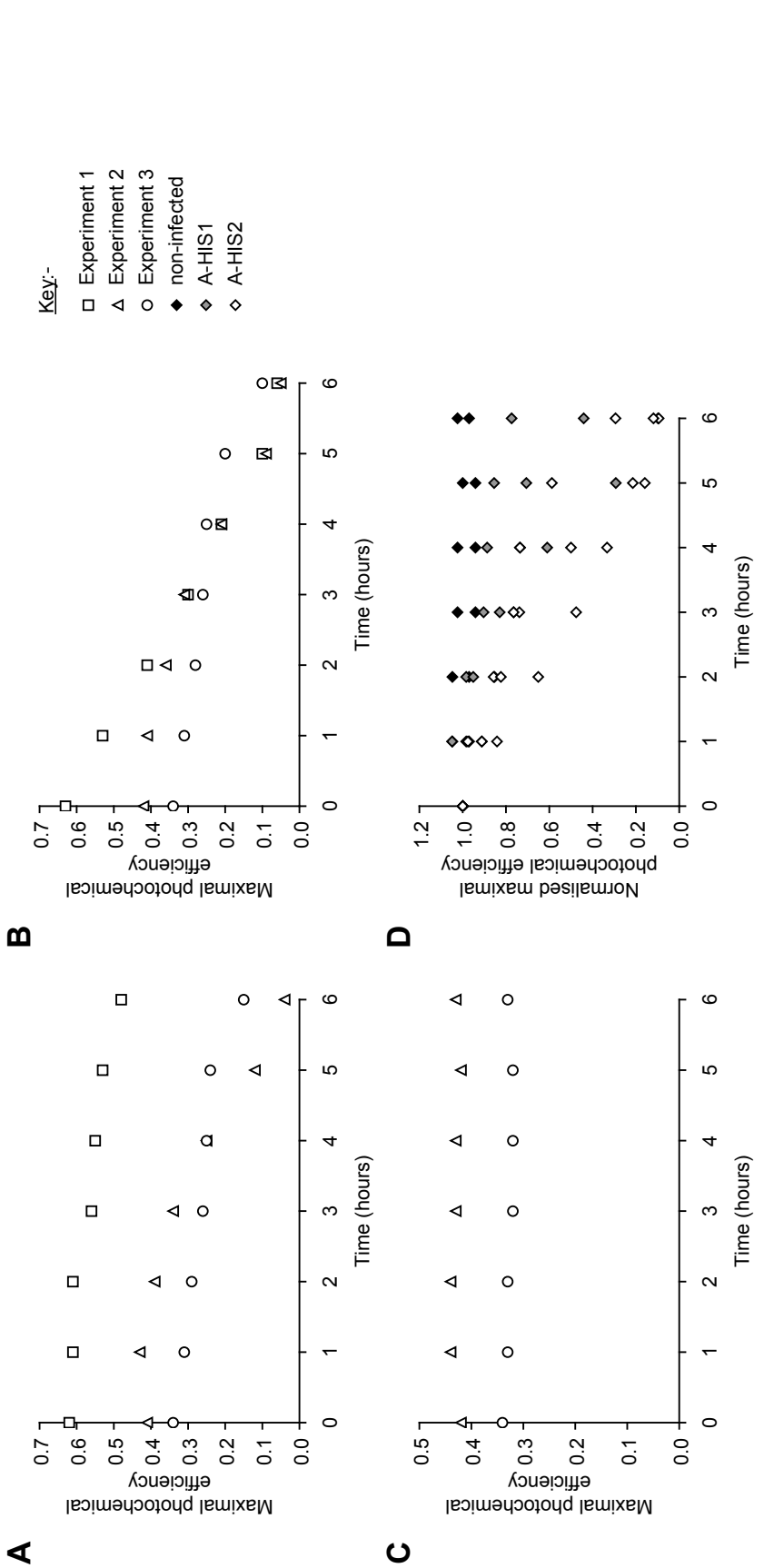


Fig. 5.6: Effect of phage on photosynthetic capacity. Measurements of maximal photochemical efficiency (quantum yield, see § 5.2.4 for details) on strain MBIC11017 inoculated at a MOI=1 with phage (A) A-HIS1, (B) A-HIS2, (C) non-infected and (D) the same data normalised by yield at zero time point and combined. Excitation at 470 nm. For (A) and (B) n=3 and for (C) n=2.

while the standard deviations are relatively large for some data points by considering the spread of the raw data (Fig. 5.6), the normalised data in Fig. 5.6(D) show that there are general trends for the gradual decrease in photochemical efficiency of the host *A. marina* for each phage, which are observed irrespective of the starting quantum yield.

Considering the replication parameters established in Chapter 3, one-step growth curves showed that A-HIS1 had a smaller burst size on average than A-HIS2 (Table 3.4, Fig. 3.5), while the latent and eclipse periods of the phages were virtually identical. A-HIS2 also has a longer tail and marginally larger genome than A-HIS1 (Table 3.3, § 4.3.1). Therefore, the differences observed in quantum yield and absorbance index (see § 5.3.3.1, Fig. 5.5), I^*_{Chld}, between the two phages is likely to be due to the cost of replication for A-HIS2 compared to A-HIS1. In terms of the amount of resources/host material which are needed during one replication cycle, A-HIS2 produces more phage in the same amount of time and therefore reduces the quantum yield and absorbance index quicker than A-HIS1 due to the drain on the cells' resources.

5.3.4 Gene expression during infection

Quantitative (real-time) PCR or qPCR, is a useful technique which can be used to quantify the abundance of phage or host transcripts during infection either absolutely, relatively or a combination of both (Clokie, 2009). In the *A. marina* and A-HIS1/2 systems, this technique was used to assess the expression of phage A-HIS1 or A-HIS2 genes during infection of exponentially growing strain MBIC11017 at an MOI of 1. The qPCR data was then interpreted with respect to growth parameters established in Chapter 3 to examine when certain genes are expressed with respect to the replication cycle.

In this study, four A-HIS1/2 genes were examined: mtDNA polymerase

γ, RNase T, the putative major capsid and the putative tail fibre protein. MtDNA polymerase γ and RNase T have never been found in a phage genome before (§ 4.3.8). Presumably, since phages A-HIS1 and A-HIS2 retain mtDNA polymerase γ in their genomes, it may replicate phage DNA early in the replication cycle, although this DNA polymerase has only been associated with mitochondrial DNA replication (Bolden et al., 1977). Concerning RNase T, it is of interest to see if and when this gene is expressed as an indication of its role in the replication of A-HIS1 and A-HIS2. RNase T is responsible for the tRNA end-turnover process although the physiological function of this remains unclear. In § 4.3.8.5, it was hypothesised RNase T may mark tRNAs for degradation amongst other possible functions. The putative major capsid and tail fibre protein were also included for purposes of comparison since they are crucial to phage replication and are expected to be expressed late in the replication cycle (Catalano, 2005). To optimise the qPCR experiments, PCR was used to determine whether the phage A-HIS1 homologues of the genes of interest were expressed (§ 5.3.4.1) before proceeding with qPCR studies with both phages A-HIS1 and A-HIS2 genes (§ 5.3.4.3).

5.3.4.1 Assessing expression by polymerase chain reaction

PCR experiments were performed to ascertain whether certain genes common to both A-HIS1 and A-HIS2 were expressed before proceeding with qPCR studies. The phage genes included in this study were: (1) mtDNA polymerase γ (§ 4.3.8.1), (2) RNase T (§ 4.3.8.5), (3) putative major capsid and (4) putative tail fibre (§ 4.3.5). Primers were designed for the phage A-HIS1 copy of these genes Table 5.3 together with the *A. marina* host gene *rpoB*, the RNA polymerase β subunit gene, which was chosen as the housekeeping gene for reference (Christensen et al., 2004).

The PCR primers were optimised using purified A-HIS1 (and A-HIS2) phage

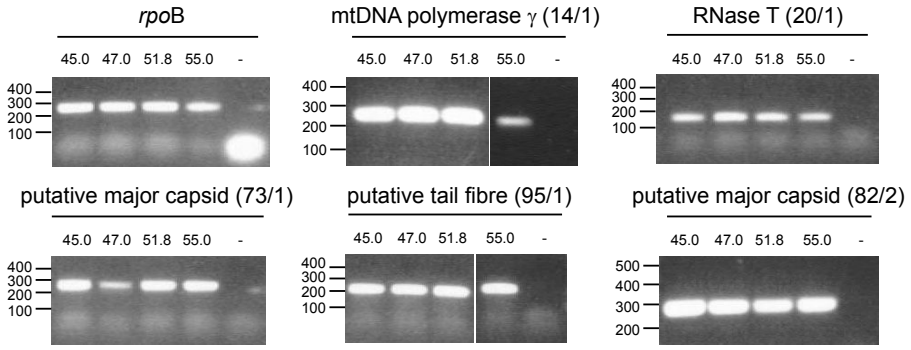


Fig. 5.7: Determination of primer annealing temperature. Each panel is labelled first by the gene that was tested for; host gene *rpoB* and phage genes A-HIS1 ORFs 14 (mtDNA polymerase), 20 (RNase T), 73 (major capsid) and 95 (putative tail fibre) and A-HIS2 ORF 82 (major capsid) (for more details, see Table 5.3). (X/Y) indicates ORF number X and phage A-HISY DNA used. Lanes are labelled by the annealing temperature used in a PCR gradient. Negative control (-) shown for each set of primers. Ladder shown in bp.

DNA and MBIC11017 host chromosomal DNA to determine appropriate annealing temperatures. PCRs performed across a gradient of 45–55 °C showed that 45 °C was a suitable annealing temperature for all phage A-HIS1 primers and host primers and PCR products corresponded to their predicted size (Table 5.3). Additionally, a PCR product (258 bp) could be amplified using the primers mcapF and mcapR for the putative major capsid gene of both phages A-HIS1 and A-HIS2 using purified phage DNA from each phage as template (last two panels, Fig. 5.7).

Following this, MBIC11017 exponential phase culture was collected pre- and post-infection with phage A-HIS1. RNA was extracted and purified by DNase treatment to remove host and phage DNA. After each treatment, to test whether samples were DNA-free, PCR was used to amplify the host gene *rpoB* and the phage A-HIS1 gene ORF 73 (major capsid).

An example of this is shown in Fig. 5.8(A)(left) where after 1 h DNase treatment, a PCR product using the *rpoB* primers could still be amplified in the 8 h sample post-infection, whereas no product is visible in the 4 h post-infection lane. In this case, the pre-infection and 8 h sample would have

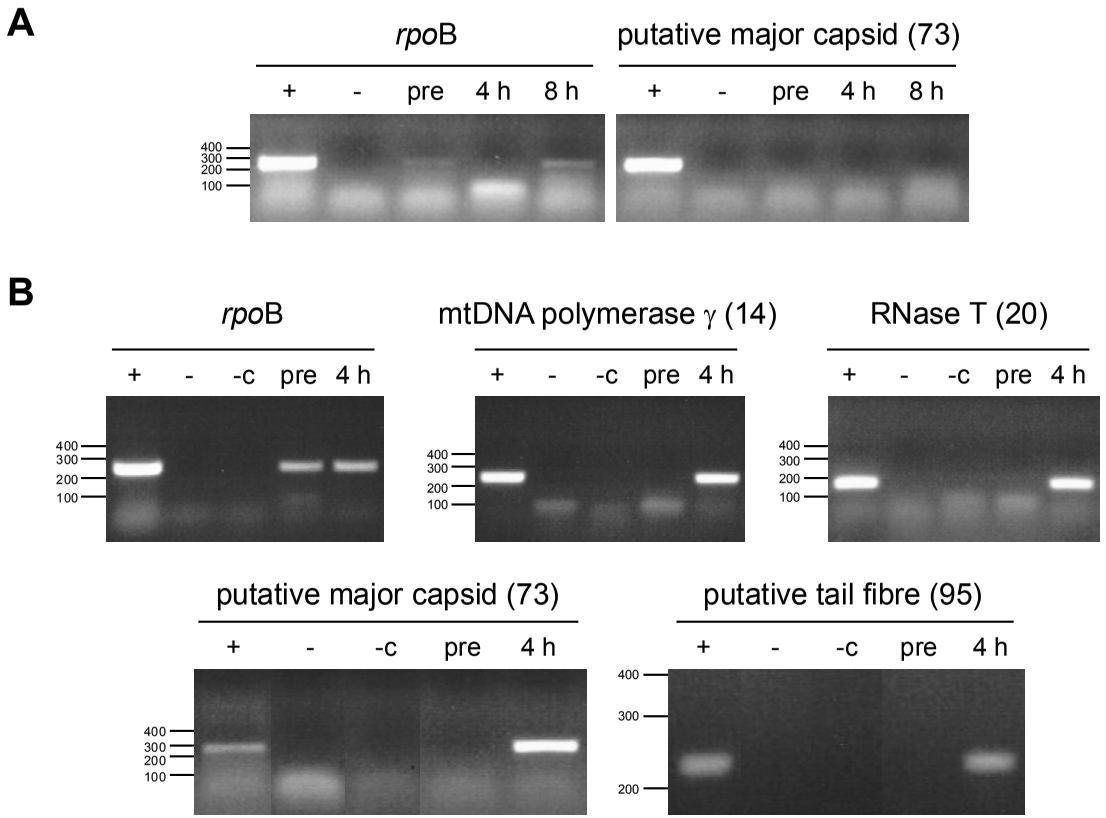


Fig. 5.8: Confirmation of gene expression by PCR. Host strain *A. marina* MBIC11017 was infected with phage A-HIS1. Each panel is labelled first by the gene that was tested for; host *rpoB*, or phage ORFs denoted by ORF number in parentheses (see details in Fig. 5.7, Table 5.3). PCR products from (A) RNA samples after 1 h DNase treatment and (B) cDNA samples. Lanes: positive control (+), negative control (-), cDNA reaction negative control (-c), pre-infection sample (pre) and post-infection samples from 4 or 8 h after infection (4 h and 8 h). Ladder shown in bp. All PCRs were performed with annealing temperature of 45 °C except for panels 73 and 95.

been treated with DNase again and tested until no PCR product could be amplified. In comparison the pre-infection, 4 and 8 h RNA samples were deemed DNA-free as no PCR product is detectable (Fig. 5.8(A)(right)). Subsequently, cDNA reactions from DNA-free RNA samples were tested which showed the host gene *rpoB* was expressed both pre- and post-infection and that the phage A-HIS1 genes ORFs 14, 20, 73 and 95 were not expressed pre-infection, but were expressed 4 h after infection (Fig. 5.8(B)).
Using an annealing temperature of 45 °C, single or multiple products could be amplified from the the pre-infection cDNA using the primers for A-HIS1 ORFs 73, and 95, indicating that the primers were susceptible to non-specfic amplification with the annealing temperature of 45 °C. Subsequently, a gradient PCR was performed from 45–65 °C and the annealing temperature was adjusted to 65 °C for A-HIS1 ORFs 73 and 95 and only single products of the correct size were detected under these conditions. These experiments indicated the phage genes of interest were expressed post-infection and could be used in qPCR experiments which is detailed in the following sections.

5.3.4.2 Intra-plate variation

In designing the qPCR experiments to be economical, intra-plate variation was assessed to maximise the use of the 96 well plates (to avoid having too many standards on each plate). Intra-plate variation was assessed by amplifying A-HIS1 ORFs 14, 20 and 30 using a fixed amount of purified phage A-HIS1 genomic DNA (5 ng, Fig. 5.9). A-HIS1 ORF 20 had the largest C_T error across the seven plates used with a standard deviation of 0.26 units. Therefore, the intra-plate variation was considered to be negligible. The equivalent averages and standard deviations in terms of copy number are included to show the effect on the C_T averages and standard deviations after conversion (Fig. 5.9(B)). Clearly, the errors are much larger with re-

gards to copy number, which can be observed by comparing the standard deviation as a percentage of the average; in the case of A-HIS1 ORF 20, this is 2.58 % and 17.92 % for the C_T and copy number data, respectively. This difference is due to the fact C_T is directly proportional to the log of the copy number and because of this, even small differences between C_T values may result in large differences in copy number with increased standard deviation.

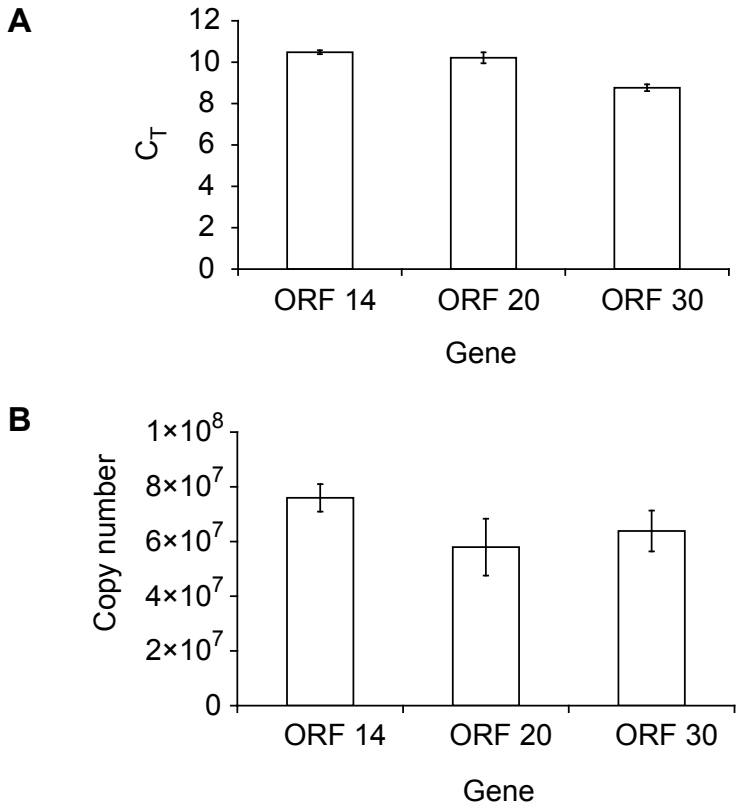


Fig. 5.9: qPCR intra-plate variation. Bar chart of (A) C_T (cycle threshold) and (B) copy number (converted from C_T values, see Equation 5.4 and Table 5.7) for A-HIS1 ORFs 14 (mtDNA polymerase γ, 20 (RNase T) and 30 (DNA pol I-like FEN). Errors are one standard deviation (n=7).

Taking this into account, the largest standard deviation from the measured C_T values was 1.20 units considering all the qPCR experimental data and therefore the replicate errors are considered to be negligible. Thus, the absolute values of gene expression are presented in terms of the average

copy number to illustrate changes in gene expression (Fig. 5.10).

5.3.4.3 *rpoB*

rpoB is a bacterial housekeeping gene (Christensen et al., 2004), which encodes the β subunit of the bacterial RNA polymerase, the enzyme involved in RNA synthesis. Therefore, this *A. marina* gene was chosen to be used as an endogenous control for relative quantitation of gene expression. However, examination of the qPCR data acquired for the *A. marina* *rpoB* showed that the copy number decreased after infection by A-HIS1 and A-HIS2 (Fig. 5.10(A/B)). This indicated that *rpoB* was not ideal as a housekeeping gene (endogenous control) for relative quantitation, as it is not constitutively expressed during infection (Pfaffl, 2006). As observed by Clokie et al. one of the major obstacles in phage transcriptomics is that it may not be possible to find a constitutively expressed gene throughout the infection cycle (Clokie, 2009). An example of this is work done with phage T4 which infects *E. coli* and completely inhibits transcription of host genes (Miller et al., 2003), which includes the housekeeping genes. Therefore, an absolute quantitation approach was adopted to assess the qPCR data.

5.3.4.4 Absolute quantitation of gene expression

Standard curves for absolute quantification using double-stranded DNA amplicons have previously been used to quantify cytokine transcript levels in mice which play a role in the immune system (Bustin, 2000; Overbergh et al., 1999). Bustin noted that the sensitivity of such standards are increased, while the amplification efficiency remains unchanged. However, it should be noted that as with previous studies, the same amount of RNA is used in each cDNA reaction, with the assumption that the cDNA reaction is equally efficient across all samples. Each pair of primers was shown to amplify evenly and efficiently (see coefficient of determination, R^2, and

amplification efficiency, E, § 5.2.11.3, Table 5.7) across a sufficient range of DNA concentrations with DNA concentration as low as the equivalent of $\sim$ 8 gene copies (§ 5.2.11.3). Importantly, the range was sufficient such that DNA concentrations from experimental samples could be interpolated from the set of standard curves (Higuchi et al., 1993).

Sample collection was performed at 2, 4, 6 and 8 h to examine the differences in gene expression between the eclipse period and the phage accumulation period (period between the eclipse and latent periods), between the phage accumulation period and the latent period and after the latent period. These times were chosen based on the phage growth parameters determined in Table 3.4, where the eclipse period extends from 0–3.25 h and the latent period from 3.25–5 h after infection at time point zero. The genes which were analysed using the absolute quantitation method were *rpoB*, mtDNA polymerase γ (A-HIS1 ORF 14 and A-HIS2 ORF 20), RNase T (A-HIS1 ORF 20 and A-HIS2 ORF 27), the putative major capsid (A-HIS1 ORF 73 and A-HIS2 ORF 82) and the putative tail fibre protein (A-HIS1 ORF 95 and A-HIS2 ORF 103), which were shown to be expressed during infection (§ 5.3.4.1).

rpoB encodes the RNA polymerase β subunit, which is involved in transcription machinery (Telenti et al., 1993). In Fig. 5.10(A/B), it can be seen that the copy number (and hence transcript abundance) of *rpoB* cDNA decreases over the duration of the experiment, to $\sim$25.6 (one standard deviation $\pm$9.3) and $\sim$11.8 (one standard deviation $\pm$2.4) copies at 8 h after infection with A-HIS1 and A-HIS2, respectively. This indicates that *rpoB* is effectively switched off and appears to be unnecessary for A-HIS1 and A-HIS2 replication. Interestingly, Allison et al. studied *rpoB* deletion mutants of the tobacco plant, which contain an *E. coli*-like *rpoB* gene in the plastid genome (Allison et al., 1996), which revealed a second transcription system in the plastids of higher plants. It is worth noting that the

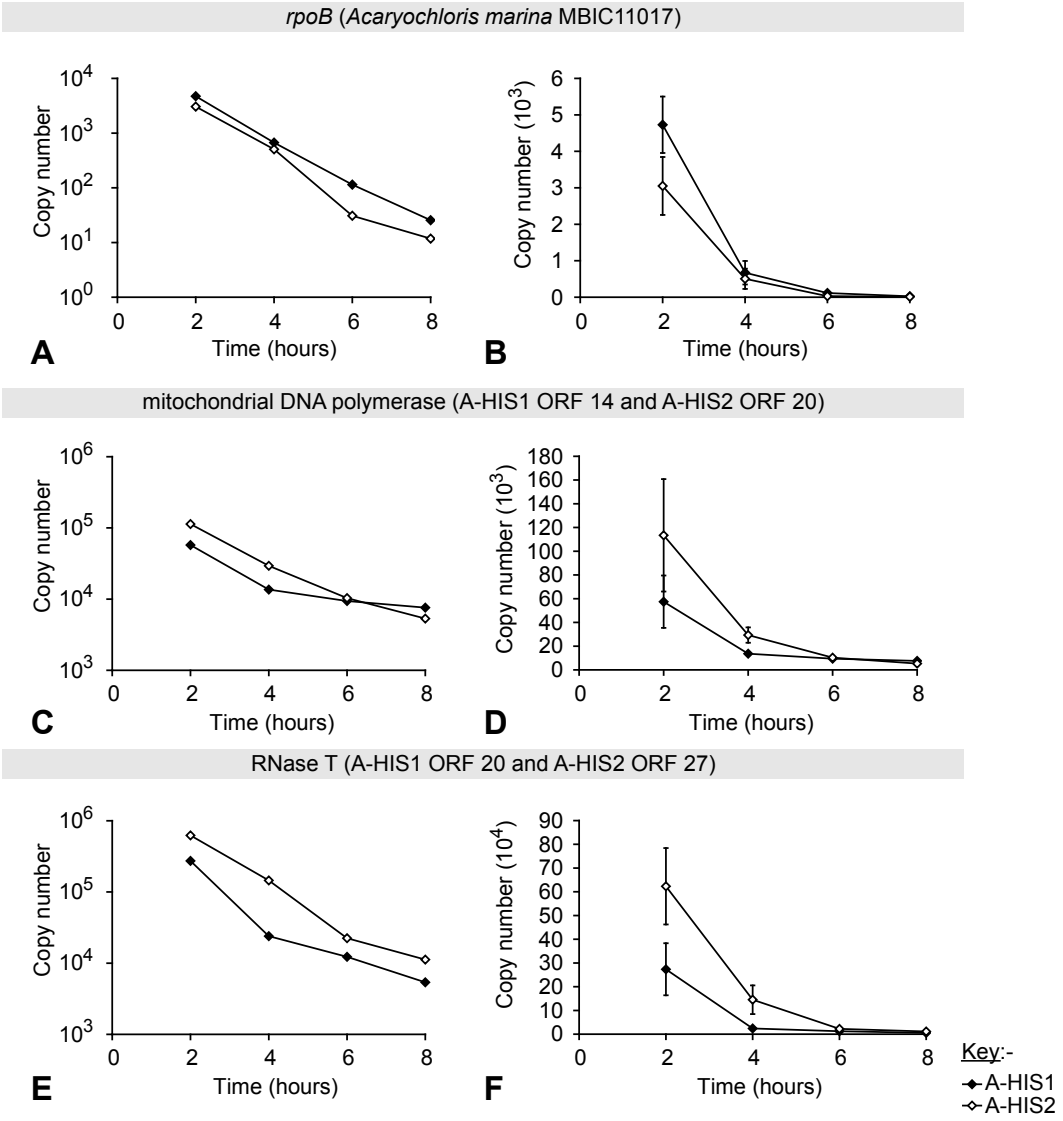


Fig. 5.10: Expression of *rpoB*, mtDNA polymerase γ and RNase T. Absolute gene copy numbers to estimate transcript abundance (converted from C_T values, see Equation 5.4 and Table 5.7) of host *A. marina* gene (A/B) *rpoB*, and *Acaryochloris* phage genes (C/D) mtDNA polymerase; A-HIS1 ORF 14 and A-HIS2 ORF 20, (E/F) RNase T; A-HIS1 ORF 20 and A-HIS2 ORF 27. Copy number with (A/C/E) logarithmic and (B/D/F) linear scale with error bars (one standard deviation). Phages infected at MOI=1 at time zero. For A-HIS1, n=2 and A-HIS2, n=3. 5 ng of cDNA was used per reaction.

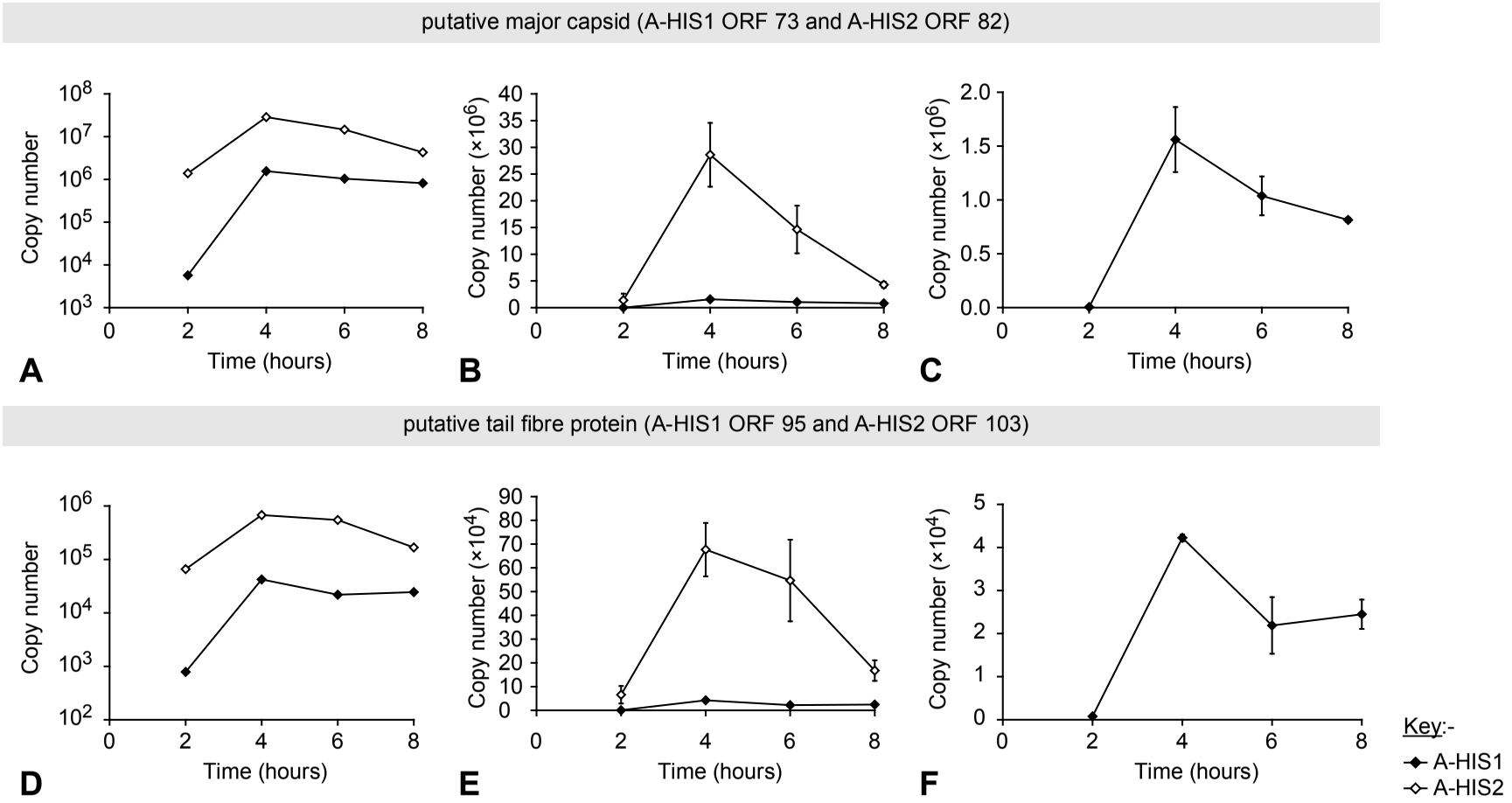


Fig. 5.11: Expression of the putative major capsid and tail fibre protein. Absolute gene copy numbers to estimate transcript abundance (converted from C_T values, see Equation 5.4 and Table 5.7) of *Acaryochloris* phage genes (A/B/C) putative major capsid; A-HIS1 ORF 73 and A-HIS2 ORF 82 and (D/E/F) putative tail fibre protein; A-HIS1 ORF 95 and A-HIS2 ORF 103. Copy number with (A/D) logarithmic and (B/C/E/F) linear scale with error bars (one standard deviation). (C/F) Graph for A-HIS1 as in (B/E) replotted for clarity. Phages infected at MOI=1 at time zero. For A-HIS1, n=2 and A-HIS2, n=3. 5 ng of cDNA was used per reaction.

plants were photosynthetically defective. However, the alternative transcription machinery was sufficient to maintain plant development. Therefore, phages A-HIS1/2 or indeed *A. marina* may possess the machinery and genes for a secondary transcription system employed by A-HIS1/2. Furthermore, given the relatively low copy numbers of *rpoB* compared to the phage genes analysed, *rpoB* is presumably switched off soon after infection. However, an analysis of *rpoB* transcript levels in uninfected cells would be required to test this hypothesis which was not included in this study.

The copy number for the phage mtDNA polymerase γ reached 57×10^3 and 113×10^3 copies at 2 h for A-HIS1 and A-HIS2, respectively, before decreasing over the remainder of the infection cycle (Fig. 5.10(C/D)). RNase T showed a similar expression profile as mtDNA polymerase for both phages. However, the copy number reached 274×10^3 and 623×10^3 for A-HIS1 and A-HIS2, respectively (Fig. 5.10(E/F)). These results indicate that mtDNA polymerase γ and RNase T are likely to be expressed early in the replication cycle, considering the eclipse period is 3.25 h for both phages (Table 3.4) and the level of expression decreases between 2 and 4 h for these genes for both phages. A-HIS2 appears to produce more transcripts than A-HIS1 (Fig. 5.10), which correlates with the fact A-HIS2 has a larger burst size than A-HIS1. The level of RNase T transcripts is also $\sim$5–6 times higher than that of mtDNA polymerase γ for both phages. Currently, the coexistence of these genes in a genome are unique to the *Acaryochloris* phages A-HIS1 and A-HIS2 and as such no comparative analysis can be performed to assess the differences in relative levels of expression of these genes in terms of their absolute quantity. Presumably, the high level of RNase T reflects the importance of this gene in the assimilation of the cell for phage replication, whatever its role (§ 4.3.8.5).

In comparison, the genes for the putative major capsid and tail fibre protein showed increased expression between 2 and 4 h after infection and

remained at steady, yet slowly decreasing levels of expression from 4 to 8 h (Fig. 5.11). The expression profiles of the putative A-HIS1 and A-HIS2 major capsids indicate expression may peak between 2 and 4 h (Fig. 5.11(A/B/C)), which is consistent with the latent period of 5 h for both phages. Significantly, A-HIS2 produces $\sim$18$\times$ more copies of the putative major capsid, which may be correlated to the difference observed in burst size. A-HIS2 also produces $\sim$16$\times$ more copies of the putative tail fibre protein than A-HIS1 (Fig. 5.11(B/E)). The expression of the putative major capsid and tail fibre structural proteins is consistent with the model of the infection cycle of phage λ (Catalano, 2005). This suggests that promoters P^1 and P^2 in A-HIS2 are likely to be late promoters in the infection cycle compared to the promoters for genes such as mtDNA polymerase γ, which are expressed before 2 h (Fig. 5.10, Fig. 5.11, § 4.3.10.1).

5.4 Conclusion

The adsorption assays indicated that while phage A-HIS1 and A-HIS2 do not infect the spontaneous phage resistant strains A-HIS1R1, A-HIS2R1 or the free-living Salton Sea strain CCMEE5410 (Miller et al., 2005), they do adsorb to these *Acaryochloris* strains to varying degrees. This indicated that there is likely to be a difference between the host receptors which the phage bind to between the wild-type MBIC11017 and strains CCMEE5410, A-HIS1R1 and A-HIS2R1. The elucidation of the cause of resistance in the phage resistant *Acaryochloris* to phages A-HIS1 and A-HIS2 requires further investigation.

PAM fluorometry showed that phages A-HIS1 and A-HIS2 reduce the maximal photochemical efficiency of the host over the infection cycle and absorbance spectroscopy showed that phage infection modifies the host strain MBIC11017 by reducing whole cell absorbance and by degrading pigments. These results indicate that while the photosynthetic capability of the host

is reduced during infection, phages A-HIS1 and A-HIS2 are able to replicate, whereas cyanophages such as S-PM2 have acquired genes involved in photosynthesis, which are thought to increase host fitness for replication (Mann, 2005; Clokie et al., 2006a).

qPCR showed that the *A. marina* gene *rpoB* is not constitutively expressed during the infection cycle. Clearly, the expression profiles for mtDNA polymerase γ, RNase T and the putative tail fibre support the previous observation that these genes are indeed expressed (Fig. 5.8), though their functionality remains to be assessed (§ 4.3.8.1, § 4.3.8.5). These qPCR results serve as a good basis for examining the *A. marina* and phage A-HIS1/2 systems, crucially showing the changes in gene expression level between the eclipse period, the phage accumulation period (between the eclipse and latent periods) and after the latent period. Future experiments could be performed between 0–2 h to resolve the expression levels of these genes during the eclipse period.

In eukaryotic mitochondria, mtDNA polymerase γ replicates mitochondrial DNA ((Bolden et al., 1977), § 4.3.8.1). It remains to be seen whether the phage mtDNA polymerases are translated and are functional. This would be the first case of a mtDNA polymerase which replicates phage DNA in a bacterium.

Chapter 6

Characterisation of *Acaryochloris marina* biofilms

6.1 Introduction

Acaryochloris spp. have been found in communal biofilms in a number of habitats, and as epibiontic colonies (Murakami et al., 2004), however, no extensive study has focussed on this aspect of their behaviour. Bacterial biofilms are ubiquitous in the environment and are studied by a diverse community of scientists varying from clinical biologists to surface chemists (Parsek and C., 2004). Interest in biofilms varies from antimicrobial resistance to more recent developments in the field of phage therapy, where phages are studied as antimicrobial agents to treat bacterial infections (Sulakvelidze et al., 2001). In light of the concurrent isolation and characterisation of phages which infect *A. marina* (Chapter 3), characterisation of host biofilms may provide further insight into the 'phage view' of its host in the natural environment with a view to designing experiments to probe phage-biofilm interactions.

This chapter investigates biofilm formation of *A. marina* with a range of microscopy techniques including optical and scanning electron microscopy (SEM), to gain insight into biofilm structure and formation. Previously,

scanning electrochemical microscopy (SECM) has been used to probe chloroplasts extracted from pea leaves in a similar SECM setup (Martin, 2007). Here, SECM is developed as a method to probe photosynthesis in small groups of live *A. marina* cells in a biofilm by photo-induced oxygen evolution at a high spatial resolution.

6.2 Materials and methods

6.2.1 Clarke electrode

Oxygen evolution was measured using a Model 10 Controller as described in (Rank Brothers Ltd., 2002). Data was recorded using a Measurement Computing USB-based 8 channel 1.2 kHz data acquisition device (USB-1208LS). For the purposes of calibration, air was bubbled through UPW and considered to be 100 % saturated. Following this, oxygen was depleted by adding sodium dithionite or gaseous nitrogen. The difference on the digital data chart between the saturation and depletion signals was then measured which corresponded to an oxygen concentration at a given temperature using ((iWorx , CB Sciences), Table 1), which was subsequently used to convert the observed changes in signal level on the digital data chart due to photo-induced oxygen. Chl *d* concentration (μg mL^{-1}) was determined by measuring A_{696} and using the specific extinction coefficient of Chl *d* in methanol together with the Beer-Lambert Law (Ritchie, 2006). Chl *a* concentrations were determined using (equation 3, (Ritchie, 2006)). Batch culture was adjusted to 5 μg mL^{-1} Chl (*a* + *d*) for experiments. Exponential phase cells (§ 2.2.3) were dark adapted for 10 min between data acquisitions and the Clarke cell was encased in foil to maintain the cell in the dark. A Zeiss Schott KL1500 was used as a light source. The contents of the Clarke electrode cell was constantly stirred using a magnetic flea and maintained at 28 °C using a water bath.

6.2.2 Preparation of biofilms

A. marina biofilms were prepared by pouring ~30 mL of exponential phase culture into Petri dishes containing autoclaved circular cover-slips. In the Material and Methods of this chapter, the word biofilm will unless otherwise stated refer to an *A. marina* biofilm grown on a glass cover-slip at room temperature under ambient light (~10 μmol photons $m^{-2}s^{-1}$). Typically, the biofilms were then transferred after 3 days into Petri dishes containing fresh autoclaved ASW using sterile tweezers.

6.2.3 Optical microscopy of biofilms

Two glass microscope slides (20 mm×14 mm×1 mm, VWR International) were taped together to form a holder for the biofilm Fig. 6.1. The biofilm was clamped between the two slides and then placed in the holder with the biofilm facing down so that the clean side of the glass cover-slip was facing the microscope lens. A drop of immersion oil was then applied to the clean side of the cover-slip and an image was acquired using the microscope. Image acquisition was performed within 2 min due to the quick evaporation of the liquid on the biofilm caused by the heat from the microscope light source. Samples were viewed using an Olympus KHC microscope.

6.2.4 Scanning electron microscopy of biofilms

The biofilms as prepared § 6.2.2 were fixed in a new Petri dish with 2.5 % glutaraldehyde for 30 min at 4 °C, at which all steps of this method were performed (method adapted from (Ortega-Morales et al., 2000)). The biofilms were then washed by transferring them to a Petri dish with UPW for 5–10 min. The wash step was repeated twice, being careful not to disturb the biofilms.

Biofilms could then be dehydrated with ethanol washes by transferring them into gradually increasing concentrations of ethanol for 5 min each

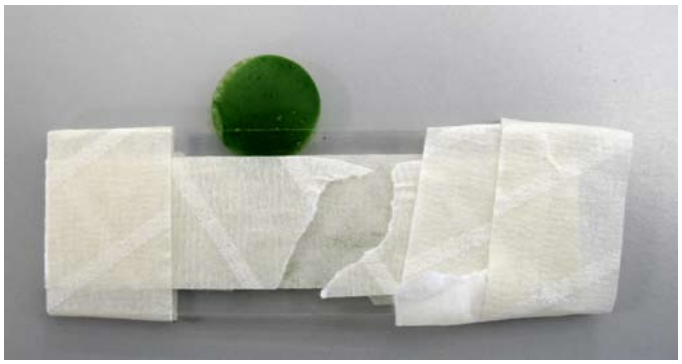


Fig. 6.1: Biofilm-holder for optical microscopy. The glass slides were taped horizontally to create a gap between the glass slides, and then taped together at both ends. The biofilm is clamped between the two glass slides and ready for image acquisition - in this example, the biofilm is face-up, and the clean-side of the cover-slip is face down.

concentration. The concentrations of ethanol used were 25, 50, 75, 95 and 100 %. Alternatively, the ethanol dehydration step was omitted. Following this, the biofilms were dried in a dessicator for 1–2 h, attached to aluminium stubs (Agar Scientific) with a strip of double-sided tape and then gold coated in a BioRad Polaron E5200 sputter coater. Images were acquired using a JEOL-JSM T330A and the software SEM:Semafore.

6.2.5 Scanning electrochemical microscopy

6.2.5.1 Setup

The SECM rig was made up of a number of components (which were produced in-house, Department of Chemistry, University of Warwick, unless otherwise stated), which were housed inside a metal Faraday cage on top of a vibration isolation stand (Fig. 6.2). The setup was based on that detailed in (Martin, 2007). The SECM was composed of a Zeiss Axiovert 40 Inverted Microscope mounted with a piezo stage equipped with micrometer-fitted x, y and z stages (Newport 462 Series ULTRAlign^{TM} Precision Integrated Crossed Roller Bearing Linear Stages), which in turn was mounted with a PI (Physik Instrumente GmbH & Co. KG) P-611 NanoCube®, which was

controlled by a PI Piezo Controller (Model E-664).

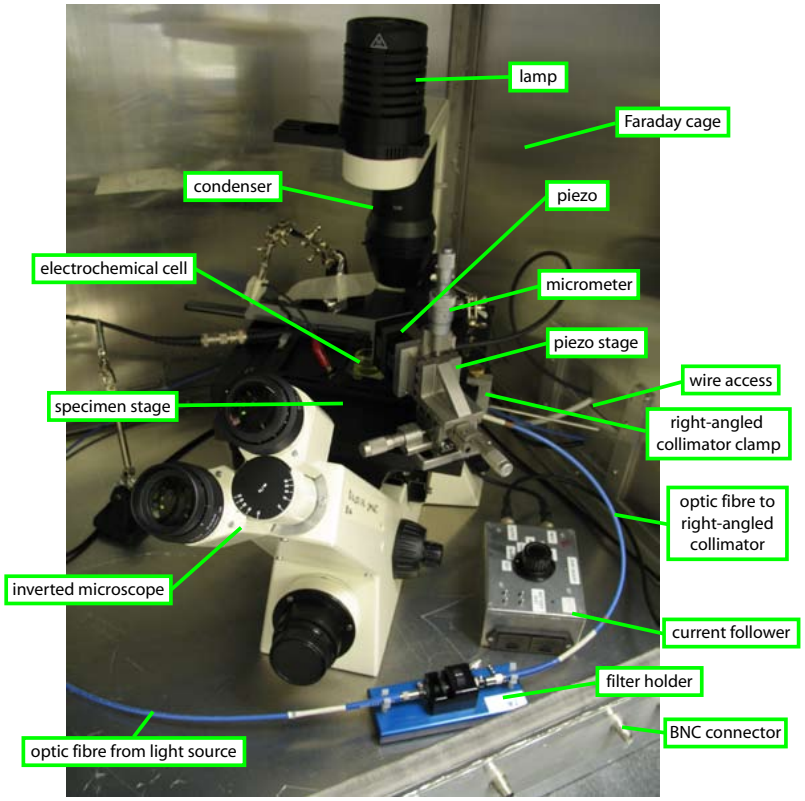


Fig. 6.2: SECM rig setup. The lens wheel and right-angled collimator are not shown here, which are positioned directly beneath the specimen stage aperture.

An aluminium ultramicroelectrode (UME) holder was then mounted onto the NanoCube Fig. 6.3. A Pt or Ag UME was then placed in the UME holder and a metal cap was placed on the end wire of the UME (working electrode, WE) which was connected to a computer-fitted data acquisition device (PCIe_6259 DAQ device, National Instruments (NI)) via BNC cables through a BNC-2110 (NI) adaptor output port. A switch was also placed in between the output port and the WE. The WE/reference electrode (RE) cable was then fed into a current follower which was connected to a BNC adaptor input port. A generalised schematic of the SECM setup is provided in Fig. 6.3. Data acquisition was performed with programs produced in-house using LABVIEW (NI) by the Warwick University Electrochemistry and Interfaces group, Department of Chemistry. Electrochemical cells were fashioned in-house by gluing a length of borosilicate tubing (~2 cm

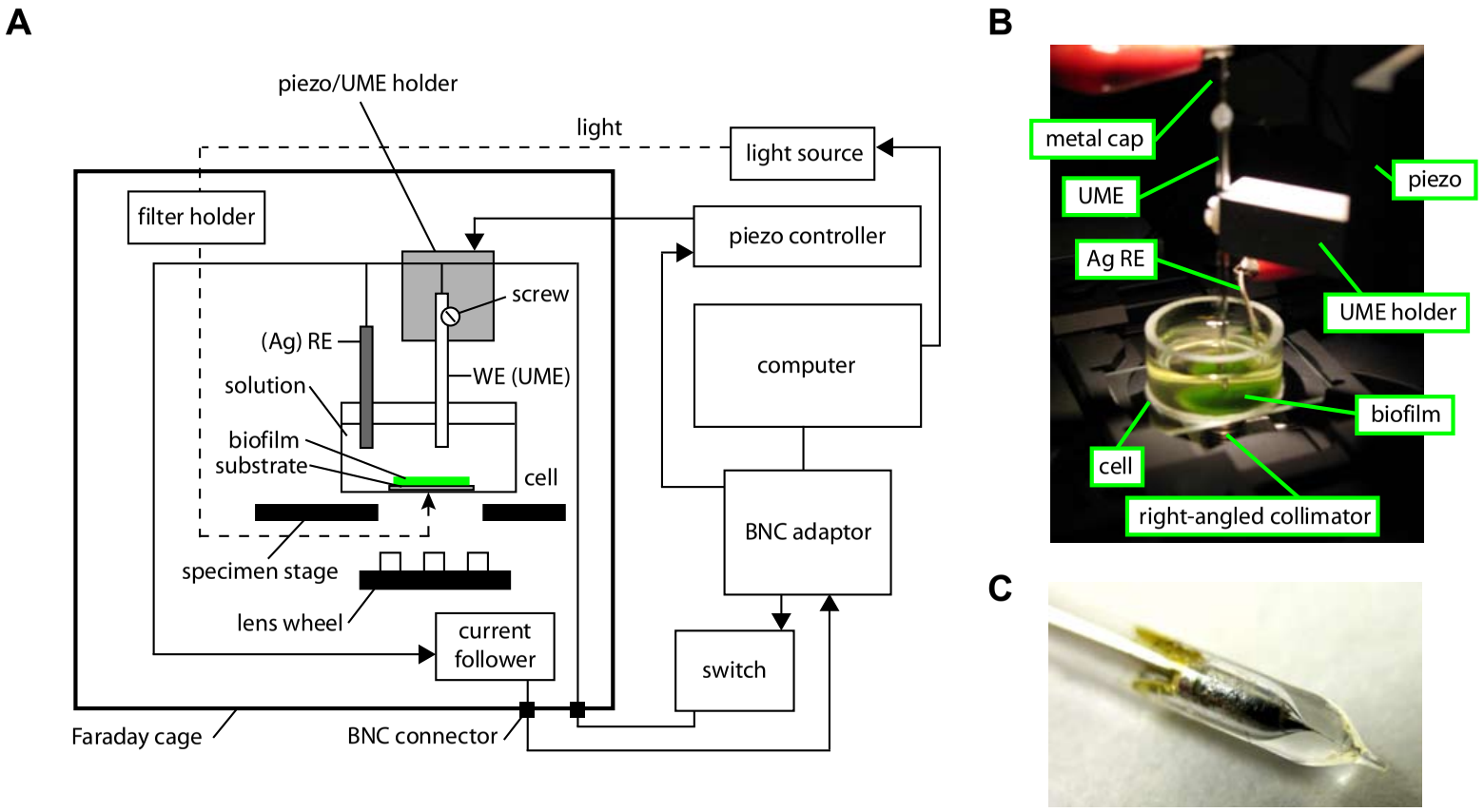


Fig. 6.3: SECM schematic and electrochemical cell detail. (A) Generalised schematic of SECM setup including glass electrochemical cell with biofilm on glass substrate in solution, working electrode (WE) and reference electrode (RE). (B) Close up of electrochemical cell setup with ultramicroelectrode (UME - the WE in this case) and Ag RE. In this photo, the right-angled (RA) collimator which directs the light can be seen through the specimen stage aperture of the inverted microscope and is flush with the base of the electrochemical cell. (C) Pt UME detail.

length/diameter) to a glass microscope slide (20 mm×14 mm×1 mm, VWR International) (Fig. 6.3(B)).

6.2.5.2 Irradiance

The sample in the electrochemical cell was illuminated using a fibre optic system from Ocean Optics, which included a HL-2000-HP-RS232 tungsten halogen light source, which was connected ultimately to a 75–90 UV right-angled collimator using two 1 mm diameter P1000-2-UV/VIS-Zferule optical fibres with a FHS-UV in-line filter holder in between. The collimator was clamped onto the side of the microscope stage so that the head of the collimator was flush with the base of the electrochemical cell. The light source was controlled via the RS232 serial port either by a LabView program or alternatively, by the accompanying Mikropack Faulhaber Motion Manager 2 software. The former option allowed automated light control with electrochemical data acquisition, whereas the latter option meant controlling the light source and data acquisition using two separate programs. The latter was only used in preliminary experiments.

Both the lenses on the lens wheel and the right-angled collimator, which are not shown in Fig. 6.2, occupied the same position beneath the stage aperture so either light was shined up towards the electrochemical cell or the underside of the cell was viewed through the inverted microscope eyepieces, but not both simultaneously (see Fig. 6.3). Light intensity was determined using a Skye Instruments Ltd. SKP 200 with lux sensor. The Faraday cage wire access portals were sealed with aluminium tape to prevent light from entering the cage so that experiments with dark conditions could be conducted.

6.2.5.3 Biofilm conditions

Biofilms were prepared as in § 6.2.2 but were not transferred to fresh ASW after an allotted time. Instead, they were transferred directly to the electrochemical cell chamber containing the appropriate buffer (Fig. 6.3(B)). For biofilm-SECM experiments, the buffer used is as described in Table 2.1, i.e. 1× ASW without trace metals (ASW^-). To this, [1]$FcTMA^+$ was added to a final concentration of 2 mM. To acquire data a 25 μm Pt UME with an RG∼14 (ratio of glass insulator diameter to metal wire diameter) was used (provided by A. L. Whitworth). For control experiments, the same buffer with 2 mM $FcTMA^+$ was used but with a clean glass cover-slip as opposed to one with a biofilm. A mark made using a permanent black marker pen was used to focus the inverted microscope to the focal plane of the glass cover-slip. The following experiments were performed with in-house programs written in LabView.

1. At the beginning of an experiment, the UME was tested by running cyclic voltammograms (CV), typically with a sweep speed of 0.01 V s^{-1}, which would give an indication of the quality of the electrode, and whether it was necessary for additional polishing or cleaning of the UME.

2. A current-time or IT program was used to measure IT transients during which species were oxidised or reduced. For oxygen reduction, the voltage was set to -0.8 V and for $FcTMA^+$ oxidation, 0.65 V was used. IT transients were measured with a sampling frequency of 15 ms with the UME at different distances from the biofilm or substrate surface, which was adjusted manually with the z-stage micrometer (see Fig. 6.2). The program also had a built in option to control the light source (§ 6.2.5.2); the right-angled collimator was alternated

[1]ferrocenylmethyltrimethylammonium hexafluorophosphate

with the microscope lenses (described in § 6.2.5.2) to acquire images of the UME/biofilm using a Canon PowerShot A620 through the inverted microscope eye-piece. Cells were allowed to dark adapt for 10 mins before data acquisition.

3. Z-approach experiments where the UME was approached towards the biofilm/substrate surface were performed typically at a rate of 0.1–1.0 μm s^{-1} via the PI Piezo Controller. The experiment was manually stopped using the program based on the output data - usually when the UME impacted a surface, the current was observed to jolt. Height and current data were recorded during these experiments.

6.3 Results and discussion

6.3.1 Oxygen evolution: Clarke electrode

An oxygen saturation curve was performed to assess the photosynthetic activity of *A. marina* cells using a Clarke electrode (Fig. 6.4). The saturation level was $\sim$32 μmol O_2 (mg Chl (a+d))$^{-1}$ h^{-1} (Fig. 6.4). The photosynthetic activity of *A. marina* has been previously measured using a similar setup, with a saturation level of 71 μmol O_2 (mg Chl (a+d))$^{-1}$ h^{-1} (Miyashita et al., 1997), which are somewhat higher than that presented, which may be due to the differences in culturing and methods of calculation involved in determining oxygen evolution.

The Clarke electrode is an established method for measuring the photosynthetic activity of photosynthetic organisms and measures 'bulk' properties of batch culture. It was therefore of interest to investigate whether these measurements could be performed on smaller groups of cells. Results from the development and assessment of techniques for this purpose are described in the following sections.

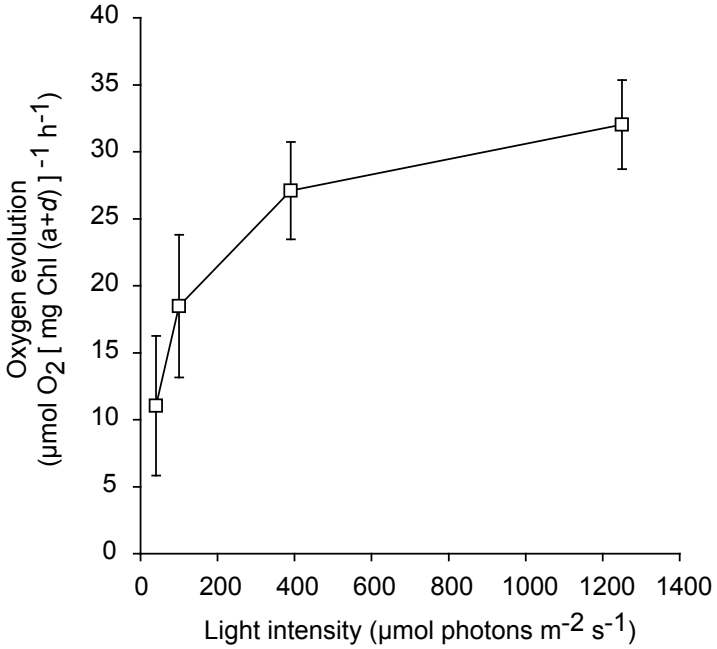


Fig. 6.4: *A. marina* oxygen saturation curve. Samples contained 5 μg mL^{-1} Chl *d* and were maintained at 28 °C. Error bars are one standard deviation.

6.3.2 Biofilm structure

Before probing biofilms with scanning electrochemical microscopy (SECM), biofilm production was optimised and biofilm structure and qualitative substrate coverage were investigated with optical and scanning electron microscopy (SEM). For the biofilm structure experiments, non-axenic culture was used (§ 3.3.1). For all other experiments, axenic culture was used. *A. marina* biofilms were found to grow successfully on glass substrates by simply allowing exponential phase bacteria to settle onto the glass substrates, which were subsequently transferred to fresh ASW (Fig. 6.5(A)). Biofilms were observed and considered to be uniform and dense on the macroscopic scale (Fig. 6.5(B)). However, on further examination with an optical microscope using a holder which allowed the biofilm-air interface to remain undisturbed (Fig. 6.1), bacterial colonies were observed to have formed on the surface of the glass substrates (Fig. 6.5(C)). These colonies were observed to merge and form larger colonies when left for two weeks, gradually producing a thicker and denser overall biofilm on the glass substrate

(Fig. 6.5(D)).

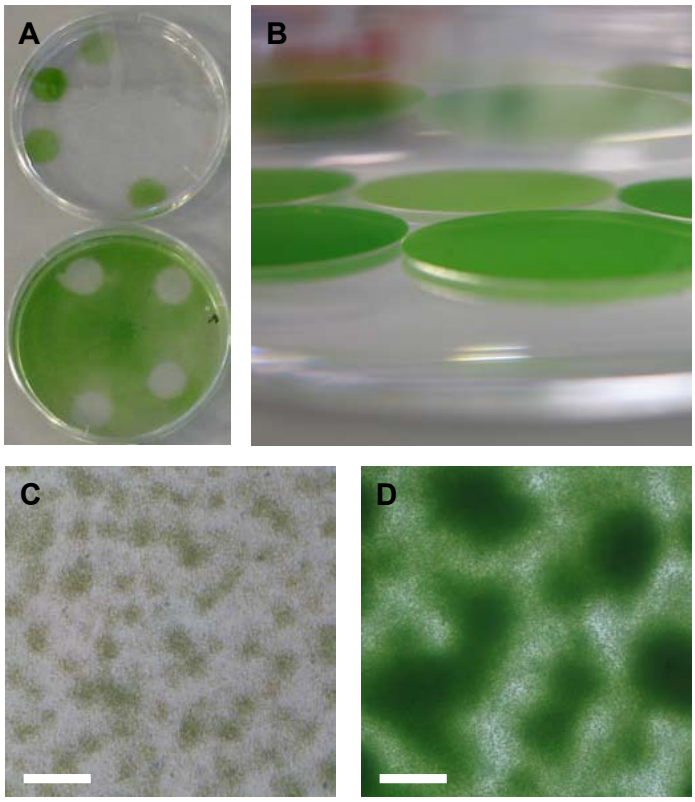


Fig. 6.5: Growing biofilims (A) *A. marina* biofilms in fresh autoclaved ASW (top) after glass substrates were left for 3 days with exponential phase culture. (B) Close up of biofilms on glass substrate in ASW. Optical microscopy images (C) 4 and (D) 12 days after start of biofilm growth. Scalebar, 50 μm.

To examine the biofilms in more detail they were subjected to dehydration treatment for SEM. At lower magnifications, the biofilms were observed to have the same uniform colony structure as observed by optical microscopy (Fig. 6.6(A,B)). At higher magnification, individual cells were observed clearly. However, there were areas of the bacterial colonies which appeared to be more homogenous and lacking cell structure (indicated by white arrows in Fig. 6.6(C,D)). These areas are thought to be extracellular polysaccharide produced by the cells which appear to form the basis of the biofilm.

At higher magnification, the cells appeared collapsed with a recessed, indented structure, contrary to the uniform ellipsoidal form expected which can be seen from longitudinal and transverse sections observed in electron

micrographs (Fig. 6.6(C,D), (Miyashita et al., 1996)). This is thought to be a result of the ethanol dehydration treatment. Indeed, SEM images of cells where the ethanol dehydration treatment was omitted showed the cell form to be significantly less resolved, each appearing to retain an extracellular coating, which is thought to be the cells' natural exopolysaccharide coating (Fig. 6.6(E)).

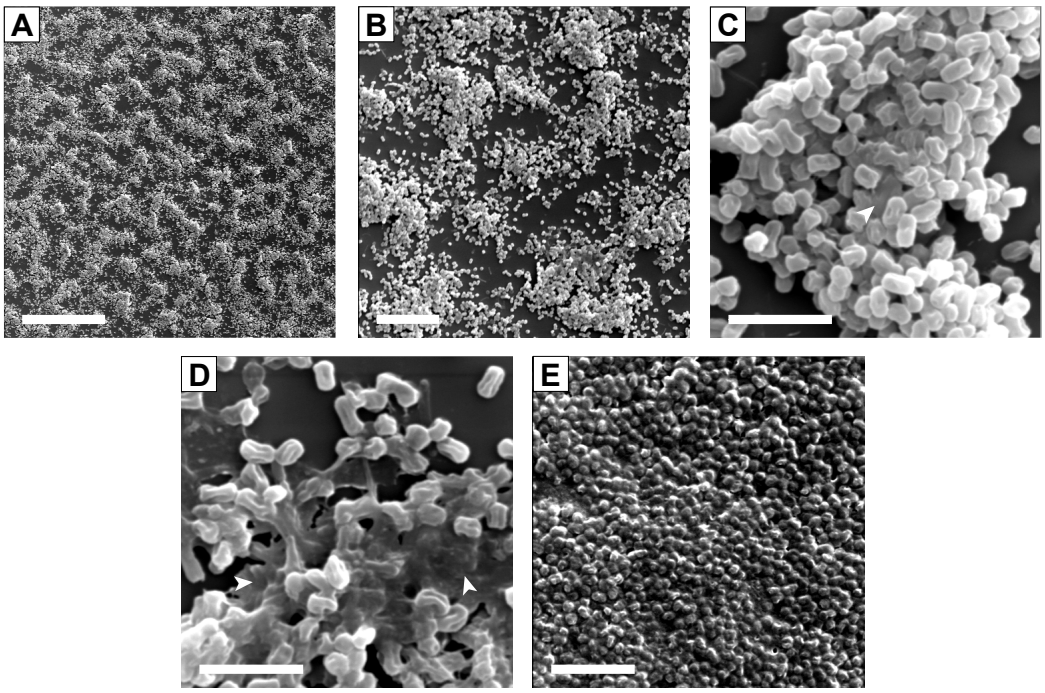


Fig. 6.6: SEM of *Acaryochloris* biofilms. (A) *A. marina* biofilm at 200×, (B) 750× and (C) 5000× magnification. (D) Biofilm polysaccharide structure and (E) non-ethanol treated biofilm. Scale bars (A) 100, (B) 20, (C,D) 5 and (E) 10 μm.

Together, the optical microscopy and SEM images show that *A. marina* is able to adhere to the glass substrate and form colonies under the laboratory conditions used (Fig. 6.6, Fig. 6.5). The microscopy data might suggest the colonies may be specifically seeded by certain cells or alternatively, that this uniform colony formation may be an artefact of the glass substrate, which is unlikely to be completely smooth and may have uniformly distributed surface features which the cells can easily tether to, thereby seeding the colonies.

However, in preliminary experiments, cells were observed to adhere within one hour to a glass substrate, although after washing the biofilm during transfer to fresh ASW to remove cells not adhered to the glass substrate, optical microscopy showed the biofilm was all but lost, leaving relatively few colonies on the substrate. Interestingly, more uniform biofilms were seen in the SECM experiments which were grown in the same way and of the same age to the SEM biofilms (3 days), and the colony substructure was not always observed (Fig. 6.7). This is thought to be due to the fact that for SECM, the biofilms were not transferred to fresh media, and care was taken to minimally disturb the biofilms which had formed before SECM data acquisition. Therefore, it is likely that the colony structures observed here were due to the cumulative effects of transferring the biofilms to fresh media and the subsequent washes before image acquisition, thereby leaving only the most strongly adhered cells on the surface.

This observation supports the idea that epibiontic *Acaryochloris* spp. may possess a mechanism for switching between a more adhesive state where cells are more likely to form colonies and a less adhesive new-habitat foraging state. If they are able to moderate their state of adhesion this would serve as an advantage for colonies when they become too dense, which ultimately limits the light available to a proportion of the cells. Indeed, *Acaryochloris* spp. have been found as single epiphytic colonies on red algae in the marine environment (Murakami et al., 2004) and these colonies are presumably surrounded and subject to turbulent fluid flow at times, of which the laboratory conditions of biofilm transfers and washes might be considered a parallel. These results also agree with the observation that *Acaryochloris* spp. occur as epibionts in their natural habitats (Larkum and Kühl, 2005).

6.3.3 Biofilm cytometry

To measure the oxygen flux produced by a group of cells in a biofilm local to the SECM probe, it was necessary to have a measure of the biofilm cell density and the amount of oxygen evolved, which would be dependent on the light intensity the cells were exposed to. Various methods were tested to obtain an accurate value for biofilm cell density. Previously, in similar studies with Chl a films (Martin, 2007), UV-visible absorption spectroscopy was used to determine a surface concentration for Chl a, however, this method was impractical for live biofilms, since the cells would ideally be minimally disturbed before SECM data acquisition.

A method combining the use of the inverted microscope and SECM was tested; the idea was to use the inverted microscope to assess the number of cells in the visible layer of the biofilm, and SECM to determine the thickness of the biofilm by measuring the distance from the UME to a chosen area of the biofilm from a fixed height and the distance from the UME to the (glass) substrate only from the same fixed height using the method of Z-approach: the height from the UME to the interface is measured together with the change in steady state current of a redox mediator as the UME is held at a fixed potential. When the UME is approached towards an inert surface the hemispherical diffusion of the redox species is limited and the corresponding current decreases until the UME contacts/crashes into the surface, upon which there is a jolt in the current and the data acquisition program is manually halted. This is the negative feedback mode of SECM and is based on the assumption that the biofilm may behave like an inert surface (Kwak and Bard, 1989). If the thickness of the biofilm could be determined, then using the dimensions of the bacteria, an approximate number of layers could be determined which could then be multiplied by the number of cells per layer and within a given volume a value for cell density in the biofilm could be calculated.

In attempts to count the cells in the visible biofilm layer, the resolution of the microscope proved insufficient, where the highest magnification was 63× for the biofilms used. This problem can be observed by considering the microscope images taken at 40× magnification on an optical microscope (Fig. 6.5) - cells on the periphery are easier to identify, but the colony centre regions of the biofilm are practically homogeneous, which is presumably due to the cells' exopolysaccharide. Indeed, *Acaryochloris* cells can routinely be counted using haemocytometric methods with optical microscopes, however, the samples are diluted and the cells are easily identifiable in these methods compared to the biofilms.

6.3.4 Biofilm thickness

Experiments to determine the thickness of the biofilms were performed with $FcTMA^+$ as the mediator which was oxidised when the UME tip was held at a positive potential of 0.65 V. Typical current-time data for the oxidation of $FcTMA^+$ in bulk solution was observed.

Interestingly, the current-time data for $FcTMA^+$ oxidation were practically identical both when the UME was in bulk solution (steady state current magnitude ∼6 nA, Fig. 6.7(A), P1) and when it was brought into contact with the biofilm, which was clearly observed using the inverted microscope, given the slight increase in the density of cells which outlined the UME tip (Fig. 6.7(A, B(P3,iii))). In fact, the current signal was only reduced when the UME pressed into the biofilm area it was aligned above as evidenced by the condensed mass of green cells around the UME tip (Fig. 6.7(A, B(P4,iv))). This experiment was repeated on 5 and 8 day old films and the same result was observed.

The reduction in oxidation current of $FcTMA^+$ from the UME tip grazing the film to being in the film was due to the reduced hemispherical diffusion to the UME tip, which is thought to be caused by the compression

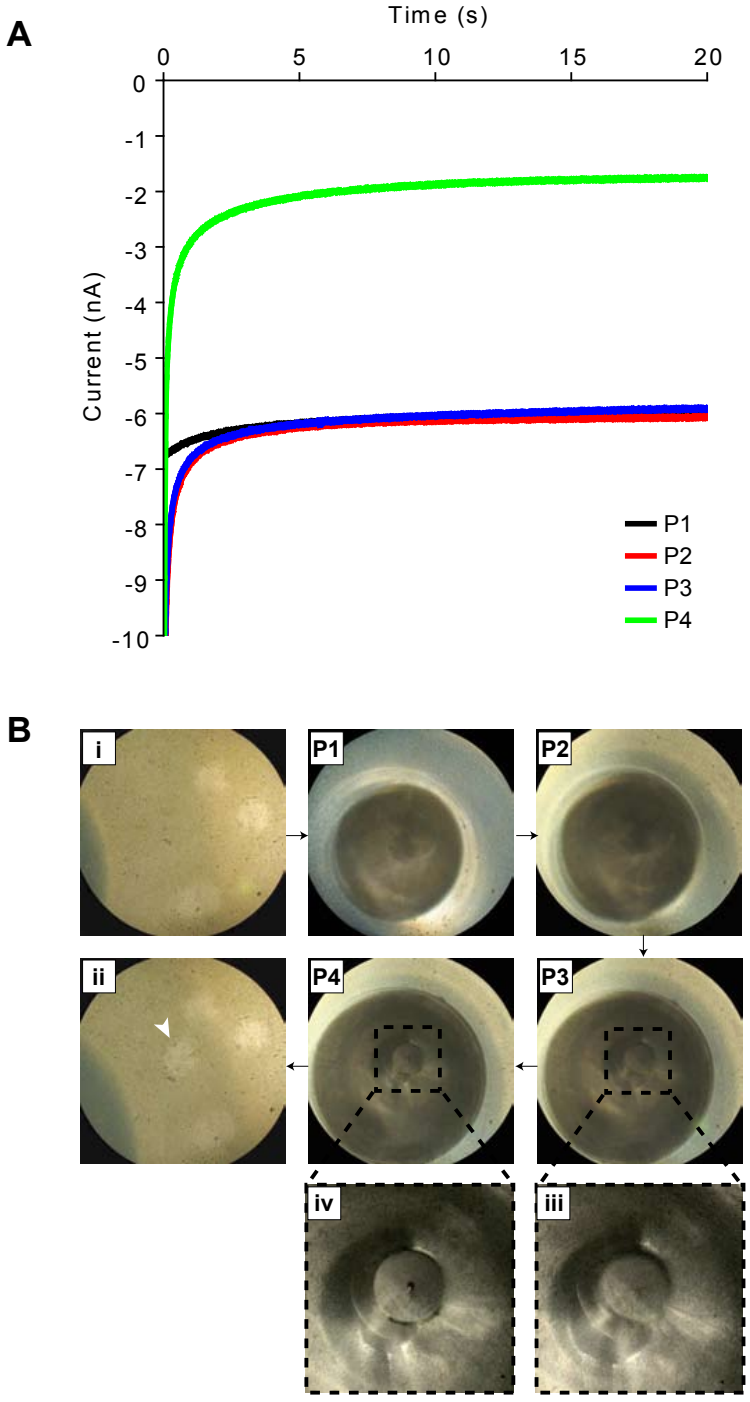


Fig. 6.7: $FcTMA^+$ oxidation above a biofilm. (A) Current-time graph from oxidation of 2 mM $FcTMA^+$ in ASW^- at different distances from a 3 day old *A. marina* biofilm. Estimated UME to biofilm distances: P1∼0.5–1 cm from film, P2∼0.2–2 mm. P3, grazing film and P4 in-film (punctured and compressed). (B) Corresponding photographs P1-P4 show UME position above biofilm taken between current-time data acquisitions. (i) and (ii) show the before and after shots of the section of biofilm over which data was acquired. White arrow in (ii) indicates the clearing of the section of biofilm after UME retraction following acquisition of P4. (iii) and (iv) show close-ups of UME-biofilm interface for images P3 and P4, respectively. Scalebar not included in photographs - for comparison UME Pt wire diameter is 25 μm. Black arrows indicate order of events.

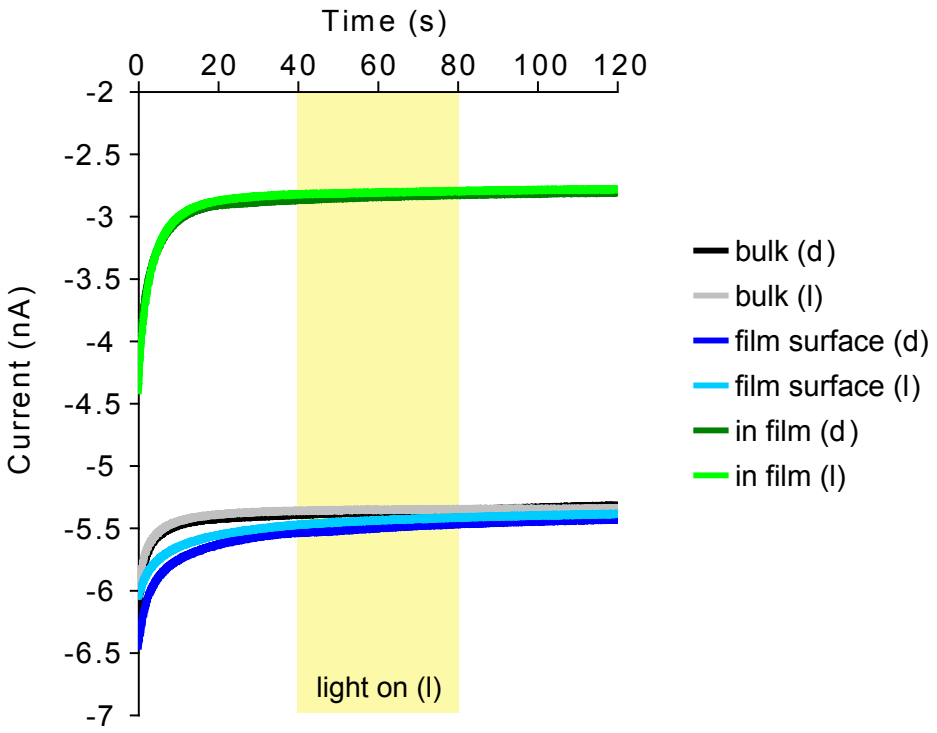


Fig. 6.8: Effect of light on $FcTMA^+$ oxidation. Current-time graph from oxidation of 2 mM $FcTMA^+$ in ASW^- at different distances from an 8 day old *A. marina* biofilm in the dark (d) or with an interval of light (l) at ~18 μmol photons $m^{-2}s^{-1}$.

and increased density of the biofilm (Fig. 6.7(A), P3 to P4). This is subtly different from negative feedback where the reduction in current is caused by the reduced distance between the UME tip and an inert surface. Surprisingly, the retraction of the UME from the 'in film' position showed the UME had permanently damaged the biofilm (Fig. 6.7(B(i,ii))). These results indicated that under dark conditions the biofilm was permeable to $FcTMA^+$, which was reduced by the compression of the biofilm. No difference was observed between current-time transients under dark conditions and those with the light source on for fixed amount of time within the transient (Fig. 6.8). Therefore, since the biofilm does not behave like a typical inert surface with respect to FcTMA, the SECM negative feedback mode could not be used to determine the thickness of the biofilm.

Interestingly, it has been shown that $FcTMA^+$ does not pass through lipid bilayer membranes and therefore has been used as a mediator for imag-

ing live cells (Wilburn et al., 2006; Ciobanu et al., 2008). Presumably, the $FcTMA^+$ does not pass through the cells themselves, but instead passes freely through the biofilm matrix. On the one hand the permeability of the biofilm serves the *A. marina* cells, since it would allow nutrients to reach cells which are not at the surface of the biofilm. On the other hand, this property does not allow the use of the negative feedback mode of SECM, where previously Amemiya et al. noted that negative feedback is observed in the application of SECM to understanding the biology of living cells, provided the cell membrane that is approached is impermeable and does not react with the redox mediator (Amemiya et al., 2006).

6.3.5 Oxygen evolution: thin biofilms

Photo-induced oxygen evolution from *Acaryochloris* biofilms was investigated using SECM. The UME tip was held at a negative potential generally between -0.7 and -0.8 V, which was determined from cyclic-voltammetric data. The steady state current from oxygen reduction was less stable than that of $FcTMA^+$ and typically varied from 4–6 nA (Fig. 6.9(A)). Oxygen reduction current-time transients showed similar behaviour to the $FcTMA^+$ oxidation data (Fig. 6.8), indicating the biofilms were permeable to oxygen as the steady state current was similar both when the UME tip was in bulk solution and when it was in contact with the film surface.

When the UME tip was far from the biofilm in bulk solution, no discernible difference was seen between the current-time transients measured in the dark and those with a fixed period of light exposure. Only when the UME was lowered directly into the biofilm, an increase of 20–30 pA in the oxygen reduction current was observed compared to the preceding dark current. This was repeatedly observed in thin biofilms from 3–8 days old when the light source was turned on (Fig. 6.9(B), Fig. 6.10(D))

On occasion current-time transients with the UME tip in the biofilm showed

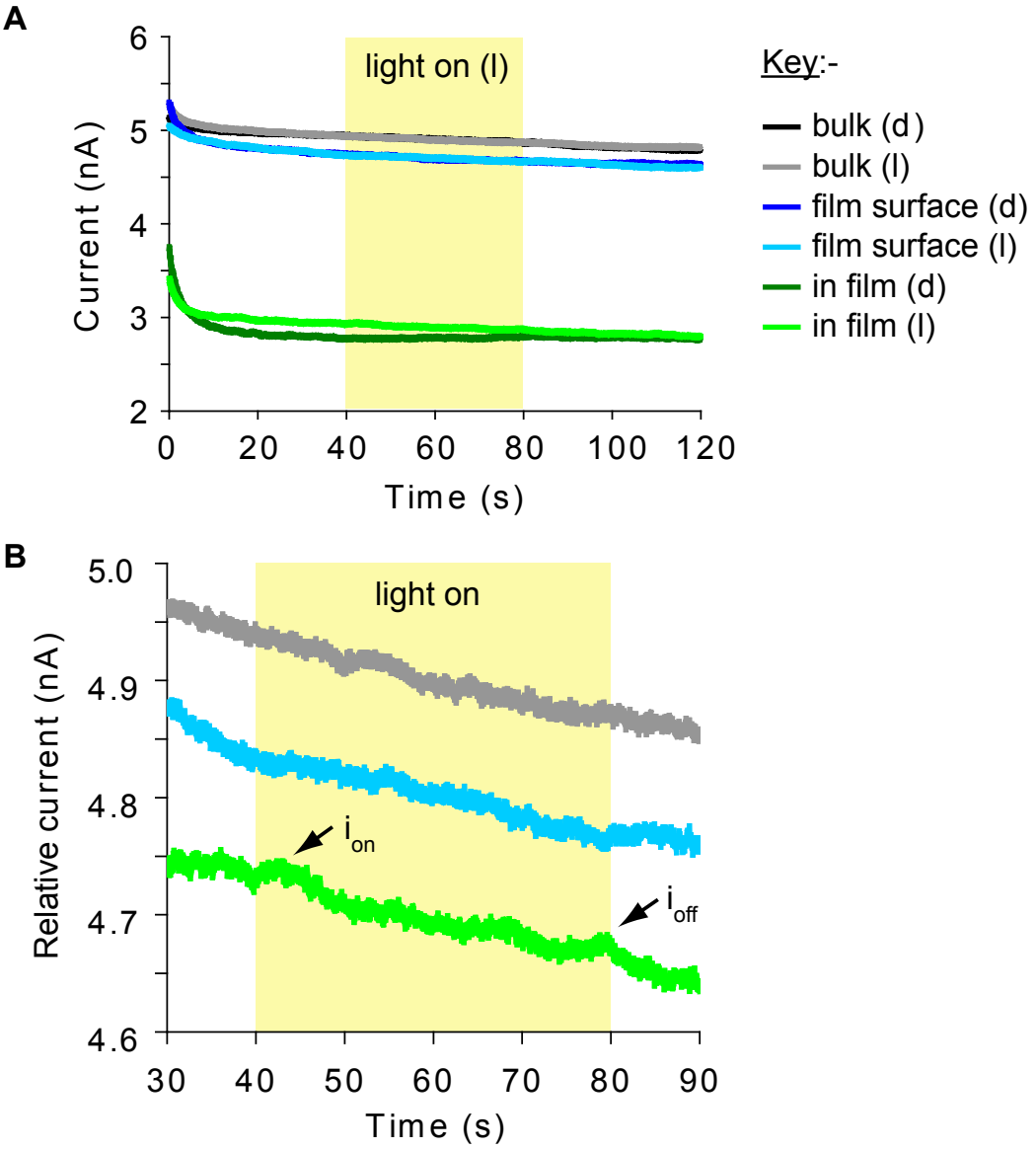


Fig. 6.9: Effect of light on oxygen reduction current. (A) Current-time graph of reduction of dissolved oxygen/photo-induced oxygen from *A. marina* cells in ASW$^-$ at different distances from an 8 day old *A. marina* biofilm in the dark (d) or with an interval of light (l) at $\sim$18 μmol photons $m^{-2}s^{-1}$. (B) Close-up of current transients from (A) with interval of light. Arrows indicate pronounced features of data collected with UME tip in biofilm.

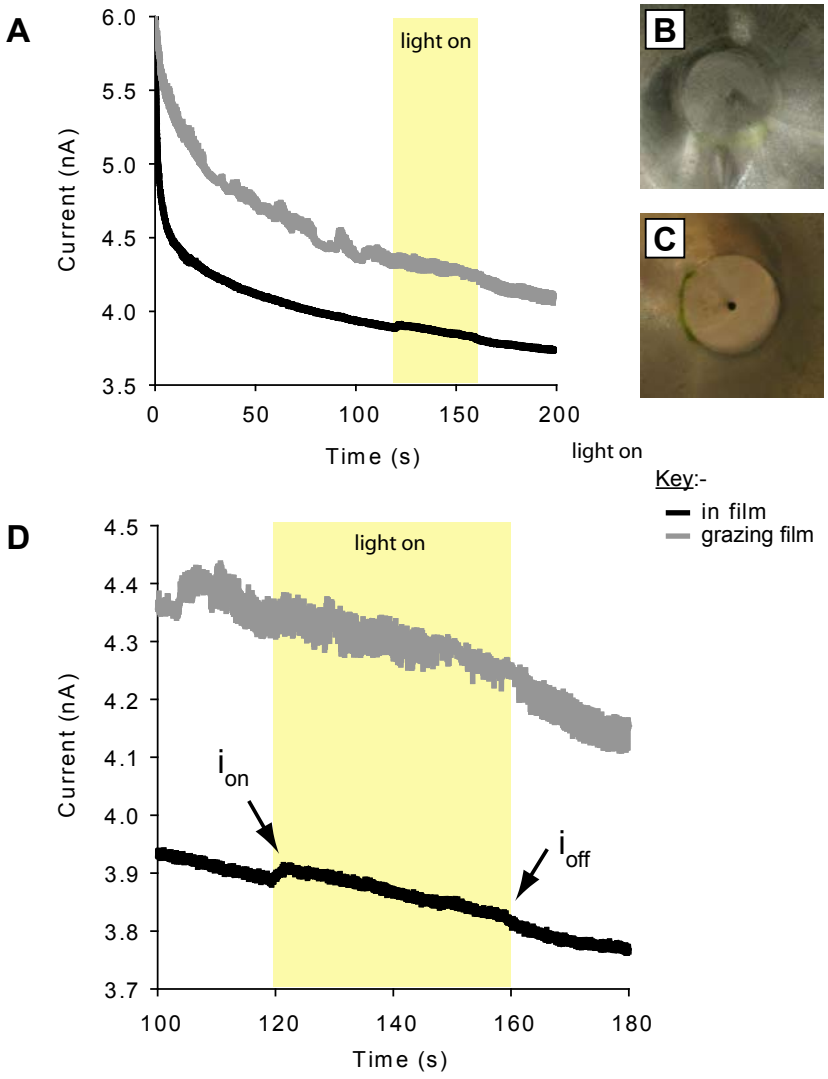


Fig. 6.10: Oxygen evolution from a biofilm. (A) Graph showing the difference between oxygen reduction current transients recorded with the UME grazing the biofilm surface and in the biofilm in the dark (d) or with an interval of light (l) at ~18 μmol photons $m^{-2}s^{-1}$. Photographs showing corresponding position of UME relative to biofilm: (B) just in film/slight compression of the biofilm surface and (C) in the film. (D) Detail of effect on oxygen reduction current signal during period of light exposure. The biofilm was 3 days old. Scalebar not included in photographs - for comparison UME Pt wire diameter is 25 μm.

a distinct discrete step increase in the oxygen reduction current corresponding to the period of light exposure as illustrated in Fig. 6.10(A,C,D) which is identified by the features i_{on} and i_{off}. However, as can be seen in Fig. 6.9(B) and Fig. 6.10(D), in attempts to replicate this result whether the UME tip was in the film or slightly compressing the film, respectively, the current-time data tended to be relatively noisy leaving i_{on} and i_{off} difficult to recognise (Fig. 6.10(D), grazing film data). The noise presumably reflects heterogeneous hemispherical diffusion to the UME tip, which may be due to non-uniform contact/blocking between the UME tip and the biofilm.

6.3.6 Oxygen evolution: thick biofilms

In light of the relative size of the noise to the data in the oxygen reduction transients (Fig. 6.9(B)) and the difficulty in reproducing a discrete photosynthetic response from the *A. marina* cells in relatively thin films (Fig. 6.10), older and thus thicker biofilms were investigated as a comparison.

In the case of thicker biofilms the UME was not easily visible compared to the thin films until it had been compressed into the film. In comparison to the thinner biofilms, the photo-induced features in current-time data, i_{on} and i_{off} for thick films as described for thin films (Fig. 6.9(B), Fig. 6.10(D)), were much more pronounced and easily recognisable which can be seen in Fig. 6.11 for 2 week old biofilms for instance, where the increase from the dark current to i_{on} was $\sim$90 pA which then continued to rise to i_{off}, where the increase from the initial dark current was $\sim$450 pA.

It has long been of interest to understand cellular activities on a single cell level to examine cellular heterogeneity compared to bulk responses of culture, for example by using techniques such as SECM. Specifically, oxygen has been shown to be a useful indicator of cellular processes such as respiration and photosynthesis (Tsionsky et al., 1997), in particular single

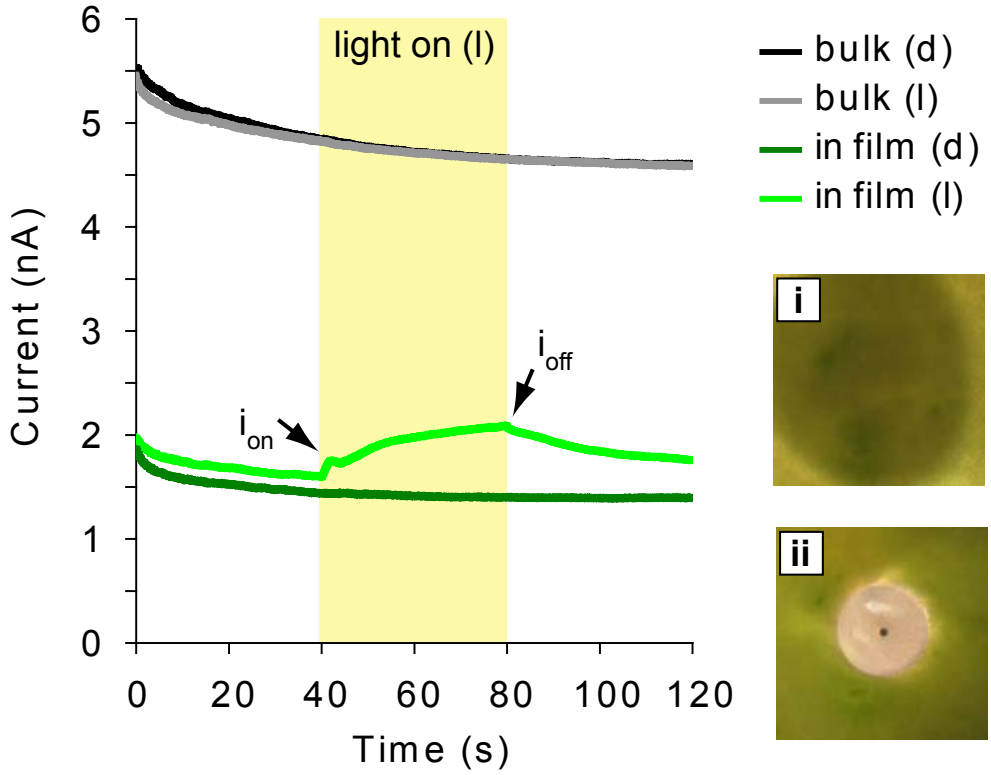


Fig. 6.11: Oxygen evolution from a thick biofilm. Graph showing the difference between oxygen reduction current transients recorded with the UME in bulk solution and in the biofilm in the dark (d) or with an interval of light (l) at ~18 μmol photons $m^{-2}s^{-1}$. Photographs showing corresponding position of UME relative to biofilm: (i) in bulk solution (ii) in the film. The biofilm was a fortnight old.

living cancer cells, and single stomatal guard cells were successfully imaged by monitoring oxygen levels (Yasukawa et al., 1998; Tsionsky et al., 1997). SECM has not been performed on cyanobacterial biofilms before and the results in this chapter serve as a basis for examining biofilms with this technique.

6.3.7 Conclusions

Biofilms of *A. marina* were produced and characterised by various microscopy techniques which showed that the artificial biofilms had an underlying colony structure which was exposed by repetitive handling and treatment of the biofilms, essentially removing the cells which were not tightly bound to the biofilms. These results provided some insight into why the epibiont *Acaryochloris* spp. have been found not only as single colonies on red algae (Murakami et al., 2004), but also as free-living species (Miller et al., 2005), with the hypothesis that *Acaryochloris* spp. may be able to switch between states: a colony/biofilm-seeding state and a foraging state. This would be a good survival strategy since invariably as an epibiontic genus, the habitats or essentially surfaces which they live on may become crowded, especially since they would be competing with other organisms in the natural environment. This may go some way to explain why *Acaryochloris* spp. have been found widespread in such a variety of habitats.

Experiments on the *A. marina* biofilms with the inverted microscope and SECM showed that although the number of cells and the thickness of the biofilms could not be accurately determined owing to the nature of the biofilms, SECM showed that the biofilms were permeable to the mediator species $FcTMA^+$ and dissolved oxygen. Moreover, an increase in the oxygen reduction current as a result of photosynthesis in *A. marina* was observed.

Chapter 7

Conclusion

7.1 *Acaryochloris* spp.: adaptability and global influence

Research on *Acaryochloris* spp. has gathered pace since the initial isolation of *A. marina* in 1996 (Miyashita et al., 1996). It has since established its reputation as a unique organism worthy of study, most notably due to its pigmentation and the resulting questions researchers have concerning its biology, ecology and evolution. Indeed, phylogenetic studies using small subunit rDNA sequences showed that the original strain, *A. marina*, was not closely related to other organisms and thus was assigned as a new taxon (Miyashita et al., 2003).

Physiologically, *Acaryochloris* spp. show many similar traits to other cyanobacteria genera and like other cyanobacteria are able to chromatically adapt (Tandeau de Marsac, 1977; Duxbury et al., 2009). In Chapter 2, it has been shown that *Acaryochloris* spp. are able to adapt their pigment composition given different light conditions which may go some way to explain their widespread occurrence in many different environments. Furthermore, the free-living strain CCMEE5410 was shown to be lacking in phycobiliproteins, pigment-protein complexes which are unique to *A. marina* (Marquardt et al., 1997; Chan et al., 2007), compared to other cyanobacteria such as *Synechococcus* spp. which contain phycobilisomes (Glazer, 1982). In

2009, the supramolecular organisation of the phycobiliproteins was finally revealed by Chen et al. in cryo-electron micrographs of ultra-thin sections of *A. marina* strain MBIC11017 (Chen et al., 2009).

Recently, Kashiyama et al. detected Chl *d* chemically from sedimentary samples obtained from many different locations as evidence that Chl *d* is globally synthesised, critically extending the range of habitats of Chl *d*-producing organisms to hemipelagic/pelagic oceans as well as lacustrine environments (Kashiyama et al., 2008). Kashiyama et al. also note importantly that photosynthesis due to Chl *d*-producing organisms should be assessed to better estimate global primary production. Concomitantly, the effect of bacteriophages which infect Chl *d*-producing organisms will also have their role to play in both global primary production and the global carbon cycle (Suttle, 2005), and should therefore not be left out of the equation.

7.2 New model system: *A. marina* and cyanophages

Bacteriophages are the most plentiful biological entities in the oceans, which have been isolated from all over the world from a whole range of environments and notably they are reported to be in their millions in every millilitre of seawater in the oceans (Suttle, 2005). Analyses of environmental metagenomic datasets have contributed to our current knowledge on the roles of phages infecting environmentally significant bacteria (DeLong et al., 2006). One reason for a metagenomic approach to study cyanophages is because they are difficult to isolate and work with. However, whilst metagenomic data offers tantalising insights, it is not possible to establish any biological information about the organisms that one has accumulated genomic data for. Thus the significance of the observed genomic diversity cannot be determined. This needs careful experimentation on isolated organisms within model systems. Additionally, to fully understand the im-

pact of cyanophages in the environment, it is crucial to acquire the genetic information for these model systems for both the cyanobacteria and the bacteriophage. Chapter 2, Chapter 3 and Chapter 4 formed the basis for studying the *Acaryochloris* system in this context.

In Chapter 2, *Acaryochloris* spp. were characterised systematically with a focus on the effects of growth irradiance on *Acaryochloris* spp. pigment composition and their ability to adapt to different light regimes as mentioned previously. In Chapter 3, phages A-HIS1 and A-HIS2 which infect *A. marina* strain MBIC11017 were isolated, which are the first of their kind. The phages were characterised and key phage parameters were established. Finally, in Chapter 4, the genomes of the *Acaryochloris* phages were characterised which showed they encode RNase T and mtDNA polymerase γ, which have never been found in a phage before. The presence of these genes also provided further insight into the origin of mtDNA polymerase γ and a link between Proteobacteria, Cyanobacteria, phage and mitochondria. Interestingly, a lack of lateral gene transfer was observed between *Acaryochloris* and phages A-HIS1 and A-HIS2, though the phages showed a high degree of synteny to each other. This and the high percentage of unidentified phage genes, indicate the novelty of the *Acaryochloris* phages and this particular phage-host system. Indeed, even when further analysis was performed on genes which have been assigned a putative function, e.g. in phylogenetics, their novelty remains abundantly clear from sequence alignments to similar sequences.

The *A. marina* phages A-HIS1 and A-HIS2 are also the first to be isolated on a cyanobacterium with a sedentary lifestyle. Chapter 5 serves as a starting point for the study of phage-host interactions on such a system, which has not been previously studied. This is largely due to the novelty of A-HIS1 and A-HIS2 and the emphasis of *Acaryochloris* research on the elucidation of the properties of the photosynthetic apparatus. In Chapter 5, the

culmination of the thesis is observed in the design and implementation of experiments to investigate the interaction between the phages and *Acaryochloris* spp. which examined the physical affinity for various hosts through adsorption experiments as well as the physiological effects of the phages on the host MBIC11017 during infection; these observations were correlated with host gene content by investigating gene expression during infection. Importantly, it was shown that the novel genes, mtDNA polymerase γ and RNase T, are both expressed early in the infection cycle. These points are discussed in more detail in the following section.

Acaryochloris spp. are naturally found living in biofilms and as such *A. marina* and phages A-HIS1 and A-HIS2 provide a new model system for the study of the interaction of phages with a biofilm. Chapter 6 details the first characterisation of *A. marina* biofilms, which grew on glass substrates and were shown to be permeable to oxygen. In Chapter 6, the characterisation of *A. marina* biofilms by microscopy indicated the *Acaryochloris* system may also be developed further to study phage-host dynamics in a biofilm, which is specifically ecologically relevant to epibiontic lifestyles of *Acaryochloris* spp. such as on red algae (Murakami et al., 2004) or beneath ascidians (Miyashita et al., 1996). It is worth taking into account that the experiments which were carried out to obtain the results presented in this thesis also serve to demonstrate the multitude of different ways the host can be studied in; in liquid culture and on agar plates (e.g. to study the property of plaque expansion, Chapter 3), to working with the host in the form of a biofilm, which has not been done before (Chapter 6).

7.3 New insights: mitochondrial DNA polymerase γ

Ever since the discovery of phage (Hankin, 1896), phage-host systems have remained enigmatic in science and the gradual elucidation of these systems requires much time, effort and one might argue, a dose of serendipity.

Notably, phages have given us much insight into different areas including the emerging field of phage therapy (Sulakvelidze et al., 2001), or their role in bacterial pathogenesis (Waldor and Mekalanos, 1996). Not least, the isolation and characterisation of the *A. marina* phages A-HIS1 and A-HIS2 (Chapter 3), clearly show that these systems can provide insight into questions which have puzzled evolutionary biologists purely based on the presence of certain genes; namely the *Acaryochloris* phages both encode mtDNA polymerase γ and RNase T homologues (Chapter 4). Indeed, such data may not have been as easily elucidated if there were not concurrently such high-throughput sequencing technologies (Hall, 2007). It is truly surprising that cyanophage are a missing link in the origin of mtDNA polymerase γ. Furthermore, the phylogenetic studies presented here suggest that a putative free-living bacterial relative of the protomitochondrion may still be extant in the oceans, which leads one to speculate whether protomitochondrial descendants may also be extant.

7.4 Future work

There are many different directions in which this work may be taken with regard to the *Acaryochloris* spp. which are now available and phages A-HIS1 and A-HIS2, which firstly would allow for a more extensive study of host range on members of the *Acaryochloris* genus. New phages may also be isolated on *Acaryochloris* spp. and sequenced to see whether properties of A-HIS1 and A-HIS2 extend to other *Acaryochloris* phages, or whether it is possible to isolate phages which show evidence of lateral gene transfer with *Acaryochloris* spp. which A-HIS1 and A-HIS2 appear to lack.

Genetic data is available for *A. marina* (Swingley et al., 2008) and in the future for strain CCMEE5410 which will allow for phylogenetic studies to compare and contrast these two distinct strains of *Acaryochloris*. The availability of the phage A-HIS1 and A-HIS2 genomes will also allow for probes

can be developed to detect the abundance and distribution of these entities in environmental samples (Fuller et al., 1998; Filée et al., 2005; Baker et al., 2006; Dorigo et al., 2004). Indeed, in Chapter 5 the PCR primers mcapF and mcapR were designed to detect the putative capsid gene of both phages A-HIS1 and A-HIS2 and so may be used in future studies.

Further phylogenetic studies may also be performed on the phage mtDNA polymerase γ and RNase T with the consideration of a molecular clock. Clearly, such studies would benefit from increased sequence data by isolating more phages or even bacteria which contain homologues of these genes. Functional and structural assays may also be developed to consider the differences between these phage genes and their eukaryotic relatives. Functional studies may also be considered to elucidate the putative function of the hypothetical ORFs which have not yet been assigned (Chapter 4). The qPCR studies considered in Chapter 5 may also be extended and developed by testing for other potential housekeeping genes for relative quantitation. Preliminary work showed that *A. marina* biofilms perished when they were incubated with phages A-HIS1 or A-HIS2 and therefore the phage-biofilm system may also be developed as a model system to study phage-host interactions in a biofilm.

APPENDIX A

Experimental data

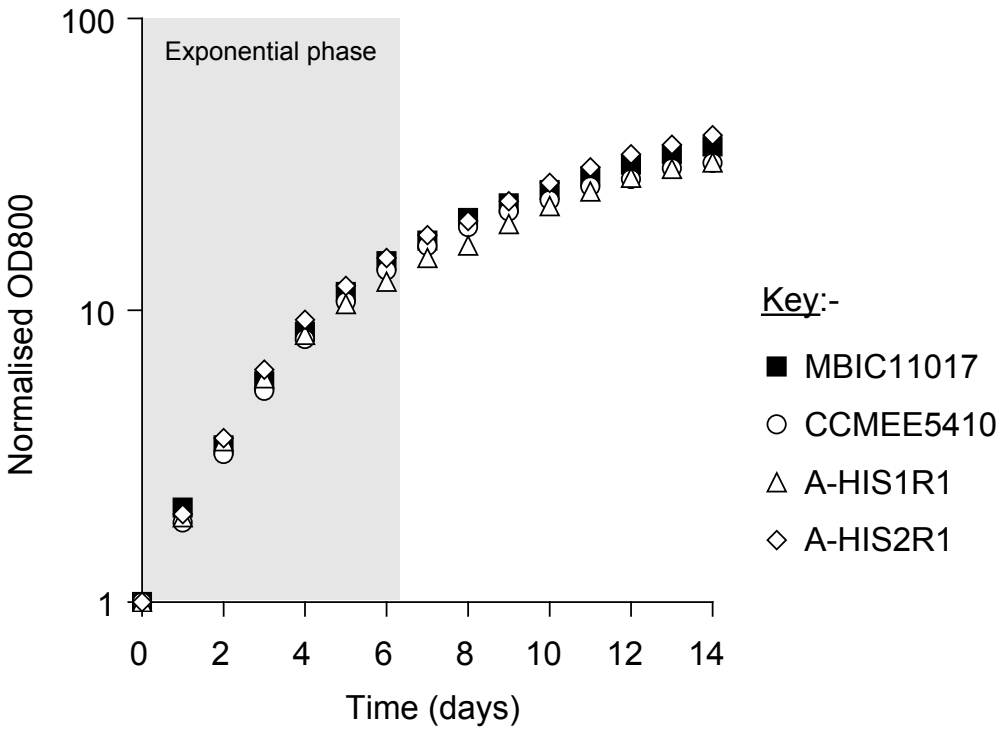


Fig. A.1: Growth curves of *Acaryochloris* strains. Cultures were grown using method detailed in § 2.2.3. For strain details see Table 3.2. Specific growth rate in exponential phase, 0.015–0.018 h^{-1} (doubling time: 38.4–45.6 h), n=2. Specific growth rate = ln(2)/doubling time.

Table A.1: A-HIS1 hypothetical proteins. Predicted ORFs where neither conserved domains not blastp hits were detected.

ORF	Start	End	Strand	Size (aa)
2	1676	2041	-	121
3	2038	2265	-	75
4	2262	2522	-	86
5	2519	2725	-	68
6	2722	3432	-	236
7	3432	3518	-	28
8	3511	3720	-	69
9	3717	3914	-	65
10	3911	4105	-	64
11	4148	4459	-	103
12	4534	4872	-	112
15	8673	8816	-	47
18	11304	11537	-	77
19	11542	12393	-	283
21	12962	13126	-	54
22	13123	13296	-	57
23	13293	13556	-	87
24	13566	14195	-	209
25	14192	14371	-	59
28	15590	16543	-	317
29	16590	16967	-	125
31	18061	18858	-	265
32	19043	19678	-	211
33	19735	20526	-	263
34	20614	20835	-	73
35	20819	20935	+	38
39	24225	24383	+	52
40	24426	24851	+	141
41	24844	25011	+	55
42	25084	26052	+	322
43	26190	26441	+	83
44	26492	26680	-	78
45	26432	26668	+	62
46	26705	27154	+	61
47	26655	26840	+	149

ORF	Start	End	Strand	Size (aa)
48	27280	27486	+	68
49	27576	28010	+	144
50	28088	28405	+	105
51	28470	29273	+	267
52	29356	29625	+	89
53	29677	29910	+	77
54	29907	30086	+	59
56	31719	31874	+	51
59	32842	33333	+	163
60	33330	33491	+	53
61	33488	34117	+	209
62	34180	34323	+	47
64	35764	36336	+	190
65	36418	36603	+	61
66	36596	37240	+	214
67	37240	37500	+	86
68	37451	37609	+	52
69	37563	37814	-	83
70	37865	38101	+	78
71	38203	38367	-	54
72	38389	38574	-	61
74	40327	41364	+	345
76	41932	42315	+	127
78	42985	43116	+	43
80	45363	45566	+	67
81	45563	45820	+	85
83	46558	46713	+	51
84	46754	47023	-	89
85	47020	47334	-	104
86	47187	47669	-	160
87	47797	47985	+	62
90	50047	50286	-	79
91	50391	50681	-	96
92	50669	50788	-	39
94	51560	52690	-	376

Table A.2: A-HIS2 hypothetical proteins. Predicted ORFs where neither conserved domains not blastp hits were detected.

ORF	Start	End	Strand	Size (aa)
2	1431	1646	-	71
3	1646	1879	-	77
4	1961	2143	-	60
5	2253	2429	-	58
6	2426	2815	-	129
7	2805	3098	-	97
8	3095	3334	-	79
9	3334	3594	-	86
10	3594	3833	-	79
11	3823	3948	-	41
12	4061	4474	-	137
13	4474	4689	-	71
14	4691	4909	-	72
15	5084	5275	-	63
16	5268	5660	-	130
17	5714	6007	-	97
18	6133	6294	+	53
19	6421	6753	+	110
21	9765	9935	-	56
23	11826	12197	-	123
24	12273	12389	-	38
25	12463	13179	-	238
26	13197	13376	-	59
28	13969	14160	-	63
29	14153	14485	-	110
30	14485	15117	-	210
33	16398	17435	-	345
36	18921	19628	-	235
37	19957	20691	-	244
38	20684	21415	-	243
39	21472	21705	-	77
40	21652	21774	+	40
41	21771	21953	-	60
42	22000	22260	-	86
43	22333	22482	-	49
44	22483	22704	+	73
45	22730	23131	+	133
46	23125	23265	+	46
47	23210	23389	+	59
48	23433	23657	+	74

ORF	Start	End	Strand	Size (aa)
49	23641	23820	+	59
50	23813	24001	+	62
51	23994	24179	+	61
52	24201	24458	+	85
53	24484	24606	-	40
54	24606	24857	+	83
56	25525	25695	+	56
58	27855	28007	+	50
60	28136	28405	+	89
61	28726	29082	+	118
62	29130	29288	+	52
63	29266	29496	+	76
64	29475	29753	+	92
65	29831	30148	+	105
66	30260	31090	+	276
67	31192	31479	+	95
70	32826	33302	+	158
71	33299	33445	+	48
72	33447	33833	+	128
74	35325	35930	+	201
75	35927	36046	-	39
76	36018	36197	+	59
77	36200	36826	+	208
78	36895	37182	+	95
79	37288	37524	+	78
80	37521	37709	-	62
81	37794	37949	-	51
84	40923	41447	+	174
86	42278	42409	+	43
88	45704	45826	-	40
89	45798	46112	+	104
91	46679	46966	-	95
92	46959	47078	-	39
93	47075	47404	-	109
94	47245	47721	-	158
95	47823	47999	+	58
96	48076	48258	+	60
98	49585	49770	+	61
100	50644	50817	-	57
102	51309	52445	-	378

Table A.3: NCBI sequence details to Fig. 4.5.

Sequence	Accession number	Organism
E1	YP_001475043	*Shewanella sediminis* HAW-EB3
E2	YP_270212	*Colwellia psychrerythraea* 34H
E3	NP_797076	*Vibrio parahaemolyticus* RIMD 2210633
E4	YP_131122	*Photobacterium profundum* SS9
E5	NP_457367	*Salmonella enterica* subsp. *enterica* serovar Typhi str. CT18
E6	NP_289353	*Escherichia coli* O157:H7 EDL933
E7	YP_404526	*Shigella dysenteriae* Sd197
E8	YP_001908650	*Erwinia tasmaniensis* Et1/99
E9	NP_670450	*Yersinia pestis* KIM
E10	AAO39143	*Photorhabdus luminescens*
F1	NP_692622	*Oceanobacillus iheyensis* HTE831
F2	NP_390083	*Bacillus subtilis* subsp. *subtilis* str. 168
F3	YP_014502	*Listeria monocytogenes* str. 4b F2365
F4	YP_186324	*Staphylococcus aureus* subsp. *aureus* COL
F5	YP_076508	*Symbiobacterium thermophilum* IAM 14863
F6	NP_625897	*Streptomyces coelicolor* A32
F7	YP_888174	*Mycobacterium smegmatis* str. MC2 155
F8	YP_119371	*Nocardia farcinica* IFM 10152
F9	YP_362679	*Xanthomonas campestris* pv. *vesicatoria* str. 85-10
F10	YP_006958	*Escherichia coli* phage T5
P1	AAR11871	*Caldicellulosiruptor saccharolyticus*
P2	ABW27165	*Acaryochloris marina* MBIC11017
P3	YP_397731	*Prochlorococcus marinus* str. MIT 9312
P4	YP_002672200	*Prochlorococcus marinus* str. MIT 9202
P5	YP_001091553	*Prochlorococcus marinus* str. MIT 9301
P6	YP_360471	*Carboxydothermus hydrogenoformans* Z-2901
P7	YP_002308942	*Candidatus Azobacteroides pseudotrichonymphae* genomovar. CFP2
P8	YP_001691651	*Finegoldia magna* ATCC 29328
P9	ZP_02094821	*Peptostreptococcus micros* ATCC 33270

Fig. A.2: Alignment of family A DNA polymerases. Sequence details (numbers in parentheses refer to those in Figure 1 in main text); NCBI (sequence name: accession number): plant; (1) *Oryza sativa* Japonica group: BAD05554, bacteria; (2) *Caldicellulosiruptor sacchorolyticus* DSM 8903: AAR11871, (3) *Carboxydothermus hydrogenoformans* Z-2901: YP_360471, (4) *Finegoldia magna* ATCC 29328: YP_001691651, cyanobacteria; (5) *Acaryochloris marina*: YP_001516479, (6) *Lyngba* sp. PCC8106: ZP_01624605, (7) *Prochlorococcus marinus* sp. CCMP1375: NP_875626, (8) *Prochlorococcus marinus* str. MIT 9312: YP_397731, (9) *Synechococcus* sp. RCC307: YP_001227925, phage; (10) *Mycobacterium smegmatis* mc2155 phage Bethlehem: AAR89764, (11) *Streptomyces* (strain Norwich stock) phage phiC31: CAA07135, (12) alpha-proteobacterium sp. JL001 phage phiJL001: AAT69504, (13) *Burkholderia cenocepacia* HI2424 phage BcepNY3: ABR10600, (14) *Escherichia coli* serotype O149:H10:F4 phage phiEcoM-GJ1: ABR68749, (15) *Synechococcus* sp. WH8109 phage Syn5: YP_001285436, (16) *Synechococcus* sp. WH7803 phage P60: NP_570330, (17) *Roseobacter* sp. SIO67 phage SIO1: AAG02598, fungi; (18) *Pichia pastoris* DSMZ 70382: Q01941, (19) *Gibberella zeae* PH-1: XP_385692, (20) *Magnaporthe grisea* 70-15: XP_364068, (21) *Botryotinia fuckeliana* B05.10: XP_001550165, (22) *Paracoccidioides brasiliensis* Pb18: EEH49594, (23) *Penicillium chrysogenum* Wisconsin 54-1255: CAP95318, (24) *Apis mellifera*: XP_395230, (25) *Anopheles gambiae* str. PEST: XP_311006, and (26) *Drosophila grimshawi*: XP_001988346. CAMERA sequences: (27) A02ORF1: JCVI_SCAF_1096627024361 (28) A05ORF9: JCVI_SCAF_1096627359208, (29) A06ORF10: JCVI_SCAF_1096627384653 and (30) A08ORF1: JCVI_SCAF_1096627376200. * denotes unambiguously aligned sequences from alignment by Filée et al. (Filee et al., 2002). G = (DNA polymerase) Gamma.

```
1_Oryza_sativa                  GRIHCSLNIN-TETGRLS-ARTPNLQNQPALEKD-RYKIRIVADYGQLELRILAHLTNCKSMLEAFKAGG-------
A_thaliana_Mus308               SVEHEVEFKL-DKNGRDV-SSD--------ADRYKINA-LLTADYSQIELRLMAHFSRDSSLISKLSQPEG------
H_sapiens_Mus308                ERIYPVSQSH-TATGRIT-FTEPNIQNVP-RDFEIKMPTILAADYSQLELRILAHLSHDRRLIQVLNTGA-------
D_melanogaster_Mus308           DRIHGQSITY-TATGRIS-MTEPNLQNVA-KEFSIQ---LLSADFCQLEMRILAHMSQDKALLEVMKSSQ-------
R_sibirica                      HRVHTTFLQTSTTTGRLS-SQEPNLQNVP-IRSSEGNQILISADYSQIELRILSHIANIDALKQAFINKD-------
L_lactis                        GKIHTRYVQDLTQTGRLS-SVDPNLQNIP-VRLEEGRKILLSSDYSQIELRVLAHISADEHLIDAFKHGA-------
T_caldophilus                   GRLHTRFNQTATATGRLS-SSDPNLQNIP-VRTPLGQRILVALDYSQIELRVLAHLSGDENLIRVFQEGK-------
E_coli                          GRVHTSYHQAVTATGRLS-STDPNLQNIP-VRNEEGRRIIVSADYSQIELRIMAHLSRDKGLLTAFAEGK-------
T_maritima                      GRIHASFNQTGTATGRLS-SSDPNLQNLP-TKSEEGKEIIVSADYSQIELRILAHLSGDENLLRAFEEGI-------
B_subtilis                      HKVHTRFNQALTQTGRLS-STDPNLQNIP-IRLEEGRKIIFAADYSQIELRVLAHISKDENLIEAFTNDM-------
S_aureus                        QRIHTRFNQTLAQTGRLS-SVDPNLQNIP-VRLEEGRKIILSADYSQIELRVLAHITQDESMKEAFINGD-------
M_leprae                        GRIHTTFNQTIATTGRLS-STEPNLQNIP-IRTNAGRQILMTADYSQIEMRIMAHLSRDEGLIEAFHTGE-------
3_C_hydrogenoformans            GKLHTTFNQTGTLTGRLA-SSEPNLQNIP-IRLELGRKLRVSADYSQIELRLLAHFSEEPKLIEAYQKGE-------
2_C_saccharolyticus             GRIHTNFIQTGTATGRLA-SAEPNLQNIP-VKYDEGKLIRIDADYSQIELRILAHISEDERLINAFKNNL-------
4_F_magna                       GKIHTTFQQTIAQTGRIS-STNPNLQNIP-IRTEEGRLIRLDADYSQIELRVLADLANDEVMLDAFKHGA-------
8_P_marinus_MIT9312             GRVHTNFNQAATATGRLS-SSNPNLQNIP-VRTEFSRRIRLSADYSQIELRILAHLADEEILINAFHKND-------
7_P_marinus_CCMP1375            GRVHTDFNQAVTATGRLS-SSNPNLQNIP-IRTEF-RRIRISADYSQIELRILAHLSGEEILQEAYKNGD-------
9_Synechococcus_RCC307          GRVHTDFNQAVTATGRLS-SSNPNLQNIP-VRTAYSRQIRLSADYSQIELRILTHLCGEEALVEAYNSGD-------
6_Lyngbya_PCC8106               QRIHTDFNQAITTTGRLS-SSNPNLQNIP-IRTEFSRQIRVSADYSQIELRILAHLSDEPILVEAYKTNR-------
5_A_marina                      HRIHTDFNQAVTATGRLS-SSNPNLQNIP-IRTEFSRQIRMAADYSQIELRILAHLSQEPRLLEAYRNGQ-------
15_Synechococcus_phage_Syn5     GRVHHSCVLN-TNTGRQA-HMRPNLAQVPSA-----SEYRVGADASGLELRCLAHYLADNSFAETVVNG--------
16_Synechococcus_phage_P60      GRLHHSCVLN-TNTGRQA-HMRPNLAQVPSA-----HEYRVGSDASGLELRCLGHYLSDGKFAEEVVNG--------
17_Roseobacter_phage_SIO1       GRVHGRVILTGAVTHRAA-HQGPNMANIPSVPHGKGAECRVGTDAAGIQLRVLAHYMNDPIYTEQVIDG--------
T3                              GKIHGSVNPNGAVTGRAT-HAFPNLAQIPGVRSPYGEQCQAGIDASGLELRCLAHFMARFDNGE-YAHEILNG----
T7                              GKIHGSVNPNGAVTGRAT-HAFPNLAQIPGVRSPYGEQCQAGIDASGLELRCLAHFMARFDNGE-YAHEILNG----
12_phage_PhiJL001               GRIHCSFNQMAADKGRLS-CEHVNMQQQP-SRDEDIAPIWAALDYSQQEPRMLVHFAEICGMPGASEAAQ-------
10_Mycophage_Bethlehem          DRVHPGINPLQARTGRMSTS-NPSAQNLP--SGD--WMVRVSVDYQAQELRVLAALANDRTMIRAFEEEA--------
11_Streptomyces_phage_phiC31    GRIHPNINTLQARTGRMSINGDFAAQTLP--SSD--WMIRGSVDFQAIEMRVLAALADVKRMKDGFVNGGSDF----
Mycobacterium_phage_D29         DRCHTFVNPLQARTSRMS-ITGIPAQTLP-ASD---WTVMASIDYQAQELRVLAALSGDRTMIQAFKDGA-------
Mycobacterium_phage_L5          DRCHTFINPLQARTSRMS-ITGIPAQTLP-SSD---WIVMASVDYQAQELRVLAALSGDRNMIEAFENGA-------
13_Burkholderia_phage_BcepNY3   GKKTADCEAILDACDRAY-LSSPEVGPILAALADAQRQLFVGADLSGIEARFSPWIAGDLELLEAFEKGV-------
14_coliphage_phiEcoMGJ1         GYLTARCGGL-TNTLRLK---HRELVNLPSIRVFGGEELRLGSDLCSLEDRCKHHFQWDPYVKKQLAPDY-------
Bacillus_phage_SPO1             NIVHPSYNIHGTVTGRLS-SNEPNAQNTP-TLFQYNFEIIVQFDYSQLELRILVCYYSRPYTIDLYRSGA-------
18_P_pastoris                   GYILPQIIPMGTITRRAVENTWLTASNAK-KNR-LGSELFVGADVDSEELWIASLIG-DSVFK-IHGGTAIGWMTLE
19_G_zeae                       GFILPQIIPMGTITRRAVENTWLTASNAK-KNR-VGSELFVGADVDSEELWIASLVG-DATFK-LHGGNAVGFMTLE
22_P_brasiliensis               GFILPQVIPMGTITRRAVENTWLTASNAK-ANR-VGSELFVGADVDSEELWIASLVG-DAQFQ-IHGGNAIGFMTLE
23_P_chrysogenum                GFILPQVIPMGTITRRAVENTWLTASNAK-ANR-VGSELFVGADVDSQELWIASLVG-DASFQ-IHGGNAIGFMCLE
20_M_grisea                     GFILPQIIPMGTITRRAVENTWLTASNAK-KNR-VGSELFVGADVDSEELWIASVVG-DATFK-LHGGNAIGFMTLE
21_B_fuckeliana                 GYILPQIIPMGTVTRRAVENTWLTASNAK-PNR-VGSELFVGADVDSEELWIASLVG-DAQFK-LHGGNAVGFMTLE
S_pombe_Gamma                   GIILPCIIPMGTVTRRAVENTWLTASNSK-KNR-LGSELFVGADVDSEELWIVALMG-DSQFR-LHGATALGMMTLE
D_melano_Gamma                  GAICPQVVACGTLTRRAMEPTWMTASNSR-PDR-LGSELLVGADVDSQELWIASVLG-DAYACGEHGATPLGWMTLS
28_D_grimshawi                  GAICPQVVVAGTLTRRAMESTWMTASNSR-ANR-IGSELLVGADVDSQELWIASVLG-DAYAHGEHGATPLGWMTLS
H_sapiens_Gamma                 GAILPQVVTAGTITRRAVEPTWLTASNAR-PDR-VGSELLVGADVDSQELWIAAVLG-DAHFAGMHGCTAFGWMTLQ
X_laevis_Gamma                  GAILAQVVSAGTITRRAVEPTWLTASNAR-ADR-VGSELLIGADVDSQELWIAAILG-EAHFAGIHGCTAFGWMTLQ
Anop                            GAIIPQVVVCGTLTRRAMEPTWMTASNAQ-RER-VGSELLVGADVDSQELWIASVLG-DA-GTGLHGGTPFGWMTLS
Apis                            GAILPLVAVCGTLTRRAVEPTWMTASNSD-TER-VGSELIVGADVDSQELWISSIIG-DAYYKKIHGATPFGWMTLI
AHIS1_ORF14                     LVIVPDLQVNGAGTRRPVDPTFLAATSSK-ATK-LGSELFVSADFNSQEMTIAGRMA-DSRAG-VLGGTPVGCLVQT
AHIS2_ORF20                     VIIIPQWECHGTLTRRAVESLWLTTTSSK-INK-LGSELFMGADFNSQEMTYAGRLG-DSIAG-ILGGTPIGVLVIC
27_A02ORF1                      TIIIPQTVPHNTATNRAGESLWLTVPDPK-PEK-IGTEVFISSDYDGQEAVVASIFA-DSYYK-VPGSTQFGHSILA
28_A05ORF9                      TVVVPQTIPHNTATNRAGEHLWLTVPDPK-PDK-IGTEVFVSSDYDGQESVVASIFA-DSEYK-IAGSTQFGHSVLA
30_A08ORF1                      NLIVPATVPHNTSTNRAGENLWLTVPDPK-YDK-IGSEIFVESDFDAQEAVVASIFA-DSYYK-VAGSTQFSHAILA
29_A06ORF10                     YMMMPEIIAHGTVTRRTVESLMVTMCSTK-GHR-IGTELIVGADFDGQELQIASIYA-DAWEGGFIGASPMTHTVLS
```

```
1_Oryza_sativa                -------DFHSRTAMNMYQVEE-KKVLLGAERRKAKMLNFSIAYGKTAVGLSWDWRAAINAPVQGSAADVAMCAMLE
A_thaliana_Mus308             -------DVFTMIAAKWTGKAE-DSVS-PHDRDQTKRLIYGILYGMGANRLAEQLRQAVNSMCQGSAADIIKIAMIN
H_sapiens_Mus308              -------DVFRSIAAEWKMIEP-ESVG-DDLRQQAKQICYGIIYGMGAKSLGEQMRQAINTIVQGSAADIVKIATVN
D_melanogaster_Mus308         -------DLFIAIAAHWNKIEE-SEVT-QDLRNSTKQVCYGIVYGMGMRSLAESLRQAVNSTIQGSAADIAKNAILK
R_sibirica                    -------DIHTQTCQIFNLQK--HELT-SEHRRKAKAINFGIIYGISAFGLAKQLRAAINAPIQGTNADIIKIAMIN
L_lactis                      -------DIHTSTAMRVFGIEKAEDVT-ANDRRNAKAVNFGVVYGISDFGLARNLRTAINSPIQGSAADILKIAMIN
T_caldophilus                 -------DIHTQTASWMFGVPP-EAVD-PLMRRAAKTVNFGVLYGMSAHRLSQELRMAFNMPVQGTAADLMKLAMVK
E_coli                        -------DIHRATAAEVFGLPL-ETVT-SEQRRSAKAINFGLIYGMSAFGLARQLRAAINAPMQGTAADIIKRAMIA
T_maritima                    -------DVHTLTASRIFNVKP-EEVT-EEMRRAGKMVNFSIIYGVTPYGLSVRLRIAINTPIQGTAADIIKLAMIE
B_subtilis                    -------DIHTKTAMDVFHVAK-DEVT-SAMRRQAKAVNFGIVYGISDYGLSQNLRTAMNTPIQGSAADIIKKAMID
S_aureus                      -------DIHTATAMKVFGVEA-DQVD-SLMRRQAKAVNFGIVYGISDYGLSQSLRTAMNTPIQGSAADIIKLAMVK
M_leprae                      -------DLHSFVASRAFGIPI-EDITPELRRR-VKAMSYGLAYGLSAYGLATQLRAALNAPIQGSAADIIKVAMIA
3_C_hydrogenoformans          -------DIHRKTASEVFGVSL-EEVT-PEMRAHAKSVNFGIVYGISDFGLGRDLRTAMNTPLQGSAADIIKLAMIN
2_C_saccharolyticus           -------DIHSQTAAEIFGVDI-SQVT-PIMRSQAKAVNFGIVYGISDYGLSRDIRIAMNSPIQGSAADIMKIAMIR
4_F_magna                     -------DIHRKT-ASEV-FKV-DKV--SLQRSNAKAVNFGIVYGIGDYSLSKDLRVALNTPIQGSAADIIKVAMVK
8_P_marinus_MIT9312           -------DIHSLT-ARLI-FEK-EEI--SDERRVGKTINFGVIYGMGIKKFARSTRAAANAPIQGSSADIIKIAMVQ
7_P_marinus_CCMP1375          -------DVHTLT-AKIL-LEK-ESI--SNERRLGKTINFGVIYGMGAQRFARSTRAAANAPIQGSSADIIKIAMVQ
9_Synechococcus_RCC307        -------DVHALT-ARLL-LDK-SEV--DEERRLGKTINFGVIYGMGAQRFARATRAAANAPIQGSSADIIKKAMVD
6_Lyngbya_PCC8106             -------DIHTVT-AQFL-FEK-DDIKDPEERRFGKTINFGVIYGMGALRFSREMRAAANAPIQGSSADIIKIAMVK
5_A_marina                    -------DVHTLT-AQLL-LEK-DQI--AEERRLAKIINFGVIYGMGPHRFAREARAAANAPIQGSSADLIKIAMVQ
15_Synechococcus_phage_Syn5   -------DIHTEL-ASIYG----------TDRKSGKGVTYCLIYGGGDHKLGSTAHAALNYLLQSAGAVICKQWLLR
16_Synechococcus_phage_P60    -------DIHTAL-AEIYG----------TDRKSGKGVTYCLIYGGGNHKLGLTAHAALNYLLQSAGAIICKLWVIR
17_Roseobacter_phage_SIO1     -------DIHTFN-KEALGR--------CKDRPTAKTFIYAFLLGAGTGMIASILHLALSVYLQGGETVIMRLAVFW
T3                            -------DIHTKNQMAAE----------LPTRDNAKTFIYGFLYGAGDEKIGQIVHAALNTLLQSAGALICKLWIIK
T7                            -------DIHTKNQIAAE----------LPTRDNAKTFIYGFLYGAGDEKIGQIVHAALNTLLQSAGALICKLWIIK
12_phage_PhiJL001             -------KYRDDPDTDNHQMM--ADMA-GISRKEAKIIFLGLCYGMGQEKLCHDLHKALNRLIQGSSADQTKAAMVA
10_Mycophage_Bethlehem        -------DLHQVTA------DAAG-----MDRKVGKMANFLTVYGGGAGKLATNAYAALNYMVQSTSRDVTASAVLR
11_Streptomyces_phage_phiC31  -------DIHMYTAQLIKGLEATK-----RDRKVFKGAGFGKVYGGGVATIARQTYAVVNYQCQSAARDVLGQAMLN
Mycobacterium_phage_D29       -------DLHLMTARAAFGDHI--TKD-DPERKYAKTVNFGRVYGGGANTVAEQTYSALNYQIQSTSRDVTCKALIR
Mycobacterium_phage_L5        -------DLHQMTADAAQ----------VPRKVGKTANFQKVYGGGAKALAEAVYSALNYMIQSSSRDVTCRALIR
13_Burkholderia_phage_BcepNY3 -------DPYKLAAAAIFQVTY-EAVT-KDQRQIGKVAMLALTYGGGAVSMAANYIAIYDKPIETLDRKILSNNMTQ
14_coliphage_phiEcoMGJ1       -------DAHLAIGVIITEQDHKDGIK-CKQRPMFKTTNYACQYGAGVPTVARSA--RFSTLCQGTGAYVFDIWIIA
Bacillus_phage_SPO1           -------DLHKAVASDAFGVAI-EEVS-KDQRTASKKIQFGIVYQESARGLSEDLRQAVNTPIQGTGSDCTLMSLIL
18_P_pastoris                 GTKNEGTDLHSKTAKILG----------ISRNEAKIFNYGRIYGAGIKFTTTLLTSRINWAIQSSGVDYLHLLIIS
19_G_zeae                     GTKSAGTDLHSRTASILG----------ITRNNAKIFNYGRIYGAGLKFAATLLTSRINWAIQSSGVDYLHLLVVA
22_P_brasiliensis             GTKAAGTDLHSKTAKILG----------ISRNDAKVFNYGRIYGAGLKFAATLMTSRINWAIQSSGVDYLHLLIIS
23_P_chrysogenum              GSKAEGTDLHSRSASILG----------ISRNDAKVFNYGRIYGAGVKFAATLMTSRINWAIQSSGVDYLHLLIVS
20_M_grisea                   GTKAAGTDLHSRTASILG----------ITRNDAKVFNYGRIYGAGLKFAGQLLTSRINWAIQSSGVDYLHLLIVS
21_B_fuckeliana               GTKAAGTDLHSRTAGILG----------ITRNDAKVFNYGRIYGAGLKFASTLMTSRINWAIQSSGVDYLHLLIIS
S_pombe_Gamma                 GKKSEGTDLHSKTAAILG----------VSRDSAKVFNYGRLYGAGLKHTTLLLTSRVNWAIQSSAVDYLHLLLVS
D_melano_Gamma                GSKSNGSDMHSITAKAVG----------ISRDHAKVINYARIYGAGQLFAETLLPTRINWVVQSGAVDFLHLMLVS
28_D_grimshawi                GSKSNGSDMHSITAKVVG----------ISRDHAKVLNYARIYGAGQQFAETLLPTRVNWVVQSGAVDFLHLMLVS
H_sapiens_Gamma               GRKSRGTDLHSKTATTVG----------ISREHAKIFNYGRIYGAGQPFAERLLTSRVNWVVQSSAVDYLHLMLVA
X_laevis_Gamma                GKKSSGTDLHSKTASTVG----------ISREHAKVFNYGRIYGAGQPFAERLLTSRVNWVVQSSAVDYLHLMLVA
Anop                          GTKATRTDMHSVTAQAVG----------ISRDHAKVLNYARIYGAGQQFAERLLPTRINWVVQSGAVDFLHLMLVC
Apis                          GTKANETDMHSVTAKAIG----------ISRNQAKIINYARIYGAGQKFAETLLPTKINWVVQSGAVDFLHLMLVS
AHIS1_ORF14                   GNKADGTDAHSQLMKYLN----------ASRQESKQPNFAILYGSGKKGLTGTFPSRCNSAIQGSGQDVSHANIVA
AHIS2_ORF20                   GNKADKTDAHNQLKSYLN----------NSRDEAKQPNFAILYGSGVKGLTGTFTSRCNWAIQSSGQDLTHSLVVA
27_A02ORF1                    GSKDDATDMHSVTAKTIG----------ISRSIAKGCNYGMLYGAGVKTLASTITMRNNWVIQSTGSAMLH----
28_A05ORF9                    GSKEDGTDMHTMTAKTIG----------ISRSIAKGCNYGMLYGSGVKTLAATITMRNNWVIQSTGSAMLHAFLTA
30_A08ORF1                    RSKDNGTDMHSMTAKAIG----------ISRAVAKGCNYGMLYGCGAKTLANTITMRNNWCIQSTGSAMLHAFMAA
29_A06ORF10                   GSKEKGTDAHTKLAKAIN----------TSRDVAKGVGFAMLYGAGVRTIANTITGRINWTIQSSGAEMLAVILTA
```

```
1_Oryza_sativa                  IEARLKGWRLLLQVHDEVILEGPTES
A_thaliana_Mus308               IYSAIAECRILLQVHDELVLEVDPSY
H_sapiens_Mus308                IQKQLETGFFILQLHDELLYEVAEED
D_melanogaster_Mus308           MEKNIERVDLVMHLHDELIFEVPTGK
R_sibirica                      LDQEIEKTRLVLQIHDELLFEAPIDE
L_lactis                        LDKALTESKLLLQVHDEIILDVPLEE
T_caldophilus                   LFPRLREARMLLQVHDELLLEAPQAG
E_coli                          VDAWLQAVRMIMQVHDELVFEVHKDD
T_maritima                      IDRELKESKMIIQVHDELVFEVPNEE
B_subtilis                      MAAKLKEARLLLQVHDELIFEAPKEE
S_aureus                        FAQKMKEAKLLLQVHDELIFEVPKSE
M_leprae                        VDKSLKQSRMLLQVHDELLFEVAIGE
3_C_hydrogenoformans            VEKELKKSRLLLSVHDELVLEVPAEE
2_C_saccharolyticus             VYKKLKKSRIILQVHDELLIESPYEE
4_F_magna                       FYNEKLKAKLILQVHDELIVDCPKDE
8_P_marinus_MIT9312             INKMNVPAKMLLQVHDELLFEVEPDS
7_P_marinus_CCMP1375            LAKLKVAANLLLQVHDELVLEVEPSV
9_Synechococcus_RCC307          LQQLLLPARLLLQVHDELVLECAPNA
6_Lyngbya_PCC8106               LRDVLYRAKLLLQVHDELVFEVPPEE
5_A_marina                      LQAALYKSRLLMQVHDELVLEVPVDE
15_Synechococcus_phage_Syn5     SYELLDDYWPLAFVHDELQISVAPSQ
16_Synechococcus_phage_P60      THELLQDYYPLAFVHDEQQLSVRADQ
17_Roseobacter_phage_SIO1       QRQAKKNFKQCAWVHDEWQTEVDEDQ
T3                              TEEMLVEFAYMAWIHDEIQVACRTEE
T7                              TEEMLVEFAYMAWVHDEIQVGCRTEE
12_phage_PhiJL001               ADEAG--YELQLQVHDEIDLTVESRE
10_Mycophage_Bethlehem          LHDAGLTPMIRLVIHDEVLASVPEAE
11_Streptomyces_phage_phiC31    MRDAGLLDYMKLPIHDEIVFSAPKAD
Mycobacterium_phage_D29         LHEAGFTPYLRLPIHDEIVASLPAEK
Mycobacterium_phage_L5          LHEAGYTPYLRLPIHDEIVASLPASE
13_Burkholderia_phage_BcepNY3   LASVLLVEQIVHHVYDEALLEVREEV
14_coliphage_phiEcoMGJ1         CNERWGDPLLSGQFHDELILQVKKGF
Bacillus_phage_SPO1             INQWLRESRICITVHDSIVLDCPKDE
18_P_pastoris                   MDYLIKDARLCITVHDEIRYLVKEED
19_G_zeae                       MDYLIRDARVAITVHDEIRYLVTEKD
22_P_brasiliensis               MDFLIRRFRLAITVHDEIRYLVRDHD
23_P_chrysogenum                MDYLIRQYRLAITVHDEIRYLVKDED
20_M_grisea                     MDYLMRRFRLAITVHDEIRYLVKDHD
21_B_fuckeliana                 MDYIIRLYRLAITVHDEIRYLVKDED
S_pombe_Gamma                   MNHLIKKARLSLTVHDEVRYLSSDKD
D_melano_Gamma                  MRWLMGSVRFCLSFHDELRYLVKEEL
28_D_grimshawi                  MRWLLGPHRFCLSFHDELRYLVKDEL
H_sapiens_Gamma                 MKWLFEEGRFCISIHDEVRYLVREED
X_laevis_Gamma                  MKWLFEAGRFCISIHDEVRYLVHSKD
Anop                            MRWLMGGSRFCLSFHDEVRYLVPERY
Apis                            MKWLMKDHRFCLSFHDEVRYLVPSKY
AHIS1_ORF14                     YAYYCDVARFTYLIHDEIISIAPHNE
AHIS2_ORF20                     YTWLTQKSRFIFSYHDEVLTIARHNA
27_A02ORF1                      --------------------------
28_A05ORF9                      MEYLTQRARFCMSVHDSVLYMCKESD
30_A08ORF1                      MEWLIKDAKFNMSVHDSILYMCPKEE
29_A06ORF10                     THWLARKAQFILSIHDELWWMVPEKY
```

Fig. A.3: Alignment of DEDD family proteins. E=Environmental (CAMERA), P=Proteobacteria, A=Alpha-proteobacteria. CAMERA accession IDs (prefices R=JCVI_READ and S= JCVI_SCAF, G=GUTLS_WRM_ELBA_READ): (E00) R_1103359526432, (E01) S_1096627366603, (E02) S_1101668051372, (E03) S_1096626851710, (E04) S_1101668085828, (E05) G_1433705844, (E06) G_1433921294, (E07) G_1433909130, (E08) S_1096627214397, (E09) S_1096627125523, (E10) R_1103359073439, (E11) R_1099972062541, (E12) R_1101751095173. NCBI accession numbers: (P00) YP_002302667, *Coxiella burnetii* CbuG_Q212, (P01) YP_001002312, *Halorhodospira halophila* SL1, (P02) YP_264915, *Psychrobacter arcticus* 273-4, (P03) YP_581127, *Psychrobacter cryohalolentis* K5, (P04) ZP_02477319, DNA polymerase III, *Haemophilus parasuis* 29755, (P05) YP_002655118, NOR51-B, (P06) ZP_01618338, marine HTCC2143, (P07) ZP_01224937, marine HTCC2207, (P08) YP_001347003, *Pseudomonas aeruginosa* PA7, (P09) NP_252218, *Pseudomonas aeruginosa* PAO1, (P10) NP_743246, *Pseudomonas putida* KT2440, (P11) YP_350248, *Pseudomonas fluorescens* Pf0-1, (P12) YP_001971382, *Stenotrophomonas maltophilia* K279a, (P13) YP_002027690, *Stenotrophomonas maltophilia* R551-3, (P14) YP_002514523, *Thioalkalivibrio* sp. HL-EbGR7, (P15) YP_002662045, HTCC5015, (A) ZP_01450509, *Rhodobacterales* bacterium HTCC2255. Sequences based on those used in (Zuo and Deutscher, 2001); *E. coli* proteins: (Z1) NP_416515 DNA exonuclease I, (Z2) AP_002464 DNA exonuclease X, (Z3) AP_000870 DNA polymerase III ϵ subunit, (Z4) ACI83035 RNase T, (Z5) AAC77122, oligoribonuclease, (Z6) ACI82335, RNase D (outgroup); *S. cerevisiae* proteins: (Z7) CAA64306 Ynt20 oligoribonuclease protein, (Z8) P53331 RNA exonuclease 1, (Z9) Q12090 RNA exonuclease 3, (Z10) Q08237 RNA exonuclease 4, (Z11) CAA96800 PAN2 poly(A) nuclease; other model organism proteins: (Z12) YP_249218 DNA exonuclease I *H. influenzae* 86-028NP, (Z13) AAC07031 DNA polymerase III epsilon subunit *A. aeolicus* VF5, (Z14) NP_389540 DNA polymerase III PolC, *B. subtilis*, (Z15) P43745 DNA polymerase III subunit ϵ, RNase T proteins: (Z16) Q9KT97 *V. cholerae*, (Z17) P57285 *B. aphidicola*, (Z18) Q9PBJ8 *X. fastidiosa*, (Z19) Q9HY82 *P. aeruginosa*, (Z20) Q4QNL5, *H. influenzae*, oligoribonucleases (ORN): (Z21) AAX88769 *H. influenzae*, (Z22) P65597 *M. tuberculosis*, (Z23) CAG38531 *H. sapiens*, poly(A) nuclease proteins (PAN2): (Z24) NP_593053 *S. pombe*, (Z25) AAA81687 *C. elegans*, (Z26) AAF59018 *D. melanogaster*, (Z27) AAH94885 *H. sapiens*, (Z28) O95453 deadenylation nucleases (DAN) *H. sapiens*, (Z29) NP_596054 CAF1 family ribonuclease *S. pombe* and (Z30) NP_506169 hypothetical protein K10C8.1 *C. elegans*. Sequences with only strain designations are gamma-proteobacterial.The amino acid positions which give rise to the DEDDh family name are marked with *; (:) and (.) are standard ClustalX notation indicating whether specific 'stronger' or 'weaker' groups are conserved (`http://portal.litbio.org/Registered/Help/clustalx/`).

```
Z10              --FT--H-G-GSHIFALDCEMC-LSEQ-------G-------LV--LTRISLVN-FD---
Z11              --DI--D-G-VENVLSLDCEMA-FTSL-------G-------YE--MIRLTIVDFFT---
Z12              -KSK--EIG--KYI-AMDCEFVGVGPE-------G-------KESALARISIVNYF----
Z29              --EM--P-Q-SGDLVAMDAEFVTLNPEENEIRPDGKTATIKPCHMSVARISCIRGQ----
Z30              --EM--P-Q-IGDLVGLDAEFVTLNEEEAELRSDGTKSTIKPSQMSVARITCVRGQ----
Z27              --EMLYP-K-M--LVGIDSEFVALQQEETEVRSDGTKSTIKPSKLSLARVSVLRGE----
Z13              --EA--P-K-SGTLVAIDAEFVSLQSELCEIDHQGIRSIIRPKRTALARISIIRGEE---
Z28              --EL--P-K-EKELVGLDAEFIKIKTDLLEF--DG-K-TVQ-MR-AVGRASCVDST----
Z1               -DGKQQS-T----FLFHDYETFGTHPA--L---DR----P--AQ--FAA-IRTD-SE--F
Z15              -MAKDFS------FFIYDYESFGVNPA--T---DR----P--AQ--FAG-IRTD-AD--F
Z5               -MSANEN-N----LIWIDLEMTGLDPE--R---DR----I--IE--IAT-LVTD-AN--L
Z24              -MPFDKQ-N----LIWIDLEMTGLDPE--K---ER----I--IE--IAT-IVTD-KN--L
Z26              GESM-AQ-R----MVWVDLEMTGLDIE--K---DQ----I--IE--MAC-LITD-SD--L
Z25              ---M-QD-E----LVWIDCEMTGLDLG--S---DK----L--IE--IAA-LVTD-AD--L
Z9               KTKL-FK-P----LVWIDCEMTGLDHV--N---DR----I--IE--ICC-IITD-GH--L
E11              -SDR-FR-G-FLPVV-VDVETGGFNSE--T---DA----L--LE--IGA-VILFINE--Q
E12              -ANR-FR-T-FLPVV-VDIETGGFNSQ--T---DA----M--LE--ISA-VILGMDD--N
P11              -AER-FR-G-FLPVV-VDVETGGFNSR--T---DA----M--LE--IAA-VILEMDA--Q
E10              -AQR-FR-G-FLPVV-VDVETGGFDSH--R---NA----L--LE--VAA-VLLAMDG--Q
P05              -RDR-FR-G-FLPVV-IDVETGGFNAD--T---DA----V--LE--IAA-TLVGMND--D
P06              -SNR-FR-G-FLPVV-IDVETGGFDCQ--T---DA----I--LE--IAASTLT-MND--E
Z22              --RR-FR-G-YLPVV-VDVETGGFNSA--T---DA----L--LE--IAA-TTVGMDE--K
P13              -ARR-FR-G-YLPVV-VDVETGGFNSA--T---DA----L--LE--IAA-TTVGMDE--K
P12              -ARR-FR-G-YLPVV-VDVETGGFNSA--T---DA----L--LE--IAA-TTVGMDE--K
P14              -AER-FR-G-YLPVV-VDVETGGFNSA--T---DA----L--LE--IAA-VTIGMDE--K
P15              -AAR-FR-G-YLPVV-VDVETGGFNSA--T---DA----L--LE--IAA-TTIAMDE--K
E01              -ADR-FR-G-FLPVV-IDVETGGFNKD--T---DA----L--LE--IAA-VFLRVEE--D
P08              -SQR-FR-G-FLPVV-VDVETGGFDSQ--R---NA----L--LE--IAA-VPIEMDE--N
P09              -SQR-FR-G-FLPVV-VDVETGGFDSQ--R---NA----L--LE--IAA-VPIEMDE--N
Z21              --QR-FR-G-YLPVV-VDVETGGFNWN--H---HA----L--LE--IAC-IPIEMDD--T
```

```
E09                -SER-FR-G-YLPVV-VDVETGGFNEA--T---DA----L--LQ--IAA-VLIDFDE--T
E07                -SER-FR-G-YLPVV-VDVETGGFNEA--T---DA----L--LQ--IAA-VLIDFDE--T
E08                -GEMILR---LLPVV-VDVETGGFNEA--T---DA----L--LQ--IAA-VLIDFDE--T
P10                -AQR-FR-G-FLPVV-VDVETGGFNPR--T---DA----L--LQ--VAA-VLLRMDP--E
P01                -RTR-FR-A-FLPVV-VDVETGGLQAE--T---DA----L--LQ--IAA-VILRAEEG-T
P07                -SRR-FR-G-FLPVV-IDIETGGFNAQ--T---DA----V--LE--IAA-VMIDMDH--Q
Z4                 --DR-FR-G-FYPVV-IDVETAGFNAK--T---DA----L--LE--IAA-ITLKMDE--Q
Z20                --NR-FR-T-FYPVV-IDIETAGFNAN--T---DA----V--LE--IAI-ITLKMDE--L
Z19                --KR-FR-G-YFPVV-IDVETAGFNAQ--T---DA----L--LE--ICA-VTLSMDE--N
Z23                --NR-FR-G-YFPVI-IDVETAGFDAK--K---DA----L--LE--LAA-ITLKMDE--N
P04_DNApolIII      -KNR-FR-G-YLPVV-VDVETGGLNAQ--T---DA----L--LE--IAA-ITLKMDD--D
A                  -RDR-FR-H-YFPVI-IDVETAGFNAQ--T---DA----L--LE--IAA-VTVRMDE--A
P02                -KER-FR-K-FLPVV-VDVETAGFNAQ--T---DA----L--LE--IAC-IPVLMNE--Q
P03                -KER-FR-K-FLPVV-VDVETAGFNAQ--T---DA----L--LE--IAC-IPVLLNE--Q
P00                -KDR-FS-G-YLPVV-VDIETGGLEPL--K---NP----L--LE--IAA-VLVEVNK--E
E05                -KNR-FR-K-YLPVV-VDIETGGFDPK--K---NA----I--LE--IAI-TLIEFEN--G
E06                -KNR-FR-K-YLPVV-VDIETGGFDPK--K---NA----I--LE--IAI-TLIEFEN--G
E04                -KNR-FR-K-YLPVV-VDIETGGFNPE--T---NA----I--LE--IAI-TLIE-EK--N
E03                -KDR-FR-K-YLPVV-VDLETGGFDPL--K---NA----I--LE--IAI-TLIE-EE--D
E02                -KEK-FR-K-FLPVV-VDIETGGFDPV--N---DA----L--LE--IAI-TLIDYDK---
Z3                 ---M-ST-AITRQIV-LDTETTGMNQI--GAHYEGHK--I--IE--IGA---VEVVN---
Z18                ---M-IN-P-NRQIV-LDTETTGMNQL--GAHYEGHC--I--IE--IGA---VELIN---
Z2                 ---M-L-----R-II--DTETCGL-Q---G----G----I--VE--IAS---VDVID---
Z16                -DNL-LD-G--TFVV-IDLEATGFD-V--E---KSE---V--ID--LAA---VRVEG---
Z17                --RL-LE-EE-TYVV-FDVETTGLSAV--Y---DT----I--IE--LAA---VKVKG---
AHIS1_ORF20        -----MR------VLGIDTETGGLDPS--K---SD----V--LT--VGV-ALADITS--N
AHIS2_ORF27        -----MI------VLGFDCETGGFNPQ--E---TS----L--LT--VGI-ALANITE--S
Z31                --QA-IE-E--ADFFAIDGEFSGIS--------DGP--SV--SA--LTN-GFDTPEE-RY
Z33                --KG-LL-Y--CDFVAIDFEFLGL---------D-----V--SA--IS--LHDTVES-RY
Z32                --KH-VD-S--AHYVSIDCEFSGLL----R---D-----F---N--LN--NKNTLQD-RY
E00_DNApolIII      ---I-IT-------V-LDTETSGLDDS--R---HE----I--LE--VAL-VSYVLSN--E
Z6                 --EA-VR-A-F-PAIALDTEFVRTRTY--Y---PQ----LG-L---IQLFDGEHL-----
                                 * *                        .

Z10                NE------VIY-E--E--LV-KP-DV--PI---V-D-YL-TR-Y---SGI-TE-EK-LTV
Z11                GK------TLF-D--H--VI-QP--I-GDI---V-D--LNSD-F---SGV-HEIDR--T-
Z12                GH------VVL-D--E--FV-KP-RE--KV---V-E-WR-TW-V---SGI-KP-EH-MK-
Z29                GPA---EGVPFMD--D--YI-ST-QE--KV---V-D-YL-TQ-F---SGI-KP-GD-LDA
Z30                GPN---EGIPFID--D--YI-ST-QE--QV---V-D-YL-TQ-Y---SGI-KP-GD-LDA
Z27                GPN---KGLPFID--D--YV-AT-DD--KV---T-D-YL-TE-Y---SGI-HP-GD-LDP
Z13                GEL---YGVPFVD--D--YV-VN-TN--HI---E-D-YL-TR-Y---SGI-LP-GD-LDP
Z28                GE----R-IIF-D--D--HVKLT-DD-VEV---V-D-YL-TK-F---SGI-VK-AD-LCP
Z1                 N-V---IGE-P-E-VF--YC-KP-AD-DYLPQ-P-G-AV--L-I---TGI--T-PQ-EAR
Z15                N-I---IGK-P-I-MF--YC-KQ-TN-DYLPA-P-E-AV--M-V---TGI--T-PQ-ECN
Z5                 N-I---LAEGP-T--I--AV-HQ-SD-EQLALMD-DWNV--R-THTASGL--V-ERVKAS
Z24                N-I---LAEGP-V--L--AV-HQ-SD-ELLNKMN-DWCQ--K-THSENGL--I-ERVKAS
Z26                N-I---LAEGP-N--L--II-KQ-PD-ELLDSMS-DWCK--E-HHGRSGL--T-KAVKES
Z25                N-I---LGDGV-D--V--VM-HA-DD-AALSGMI-DVVA--E-MHSRSGL--I-DEVKAS
Z9                 APVK--AADGQGDSHYESVI-HY-GP-EVMNKMN-EWCI--E-HHGNSGL--T-AKVLAS
E11                GRLTPGF--NL-F--H--SI-IPFQG-ANIE--Q-A-AL--D-I---TGI-DP-NDPYRA
E12                GIL--GIEQTY-F--H--RI-IPFEG-LKLE--E-A-AL--K-F---TGI-DP-FHPLRV
P11                GNI--QIKETH-S--K--NI-DPFPG-AVVE--K-A-AL--E-F---TGI-DI-YDPERM
E10                GNL--QIKESY-S--K--NI-EPFPG-AIVE--E-S-AL--E-F---TGI-DI-YDPERM
P05                GIL--TPKQTL-C--H--NV-EPFEG-ANLE--P-S-AL--E-F---TGI-DP-ESPLRD
P06                GII--CIDKTF-S--F--NV-APFEG-ANLE--P-A-AL--E-F---TGI-DP-YHPFRE
Z22                GFL--FPEHTY-F--F--RI-EPFEG-ANIE--P-A-AL--E-F---TGI-KL-DHPLRM
P13                GFL--FPEHTY-F--F--RI-EPFEG-ANIE--P-A-AL--E-F---TGI-KL-DHPLRM
P12                GFL--FPEHTY-F--F--RI-EPFEG-ANIE--P-A-AL--E-F---TGI-KL-DHPLRM
P14                GFL--FPEHTY-F--H--RV-EPFEG-ANIE--P-A-AL--E-F---TGI-KL-DHPLRM
P15                GFV--YPDHTY-F--F--RV-EPFEG-ANVE--P-A-AL--E-F---TGI-KL-DHPLRM
E01                GTL--AIDDSL-T--L--HV-LPFEG-ANIE--Q-A-AL--D-F---TGI-QP-YNPLRG
P08                GLL--YPGQTA-S--A--HV-VPAEG-LEID--P-K-SL--E-V---TGI-IL-DHPFRL
P09                GLL--YPGQTA-S--A--HV-VPAEG-LEID--P-K-SL--E-V---TGI-IL-DHPFRL
Z21                GRL--YPGEVT-S--T--HV-IPAPG-TDID--P-K-SL--E-V---TGI-IL-DHPFRF
E09                GQL--YCTETV-S--C--HV-NPFEG-ANLD--P-K-SM--E-V---NGI-DV-DHPFRL
E07                GQL--YCTETV-S--C--HV-NPFEG-ANLD--P-K-SM--E-V---NGI-DV-DHPFRL
E08                GQL--YCTETV-S--C--HV-NPFEG-ANLD--P-K-SM--E-V---NGI-DV-DHPFRL
P10                GQL--YRHQTL-S--F--HV-QPFEG-ANLD--P-K-SL--E-V---NGI-NP-WHPLRI
P01                GRL--YPAETH-T--C--HV-QPFEG-ARID--P-K-AL--E-L---NGI-DP-EHPLRM
P07                GRL--RVGETI-S--T--HV-EPFQG-ANIE--P-K-SL--E-I---TGI-DL-NNPLRG
Z4                 GWL--MPDTTL-H--F--HV-EPFVG-ANLQ--P-E-AL--A-F---NGI-DP-NDPDRG
Z20                GWL--HKEDTL-H--F--HI-EPFKG-SIIN--S-D-AI--A-F---NKI-DP-FNPLRG
Z19                GDL--HPASTI-H--F--HV-EPFEG-ANLE--K-A-AL--E-F---TGIYDP-FSPLRG
Z23                GYL--HPDQKC-H--F--HI-KPFEG-ANIN--P-E-SL--K-F---NGI-DI-HNPLRG
P04_DNApolIII      GYL--HLDEKF-H--R--HV-KPFEN-ANIN--P-E-AL--K-V---NGI-DV-DNAERG
A                  GML--HPDQEI-H--Y--HI-NPFEG-ANIE--P-A-AI--E-F---NGI-KV-DCALRG
P02                GVF--YPGEAL-N--A--HI-EPFEG-ANLE--P-S-AL--A-F---TGI-DP-NNPLRK
P03                GVF--YPGEAF-N--A--HI-EPFEG-ANLE--P-S-AL--A-F---TGI-DP-NNPLRK
P00                GQL--CQGELF-A--C--HV-LPFEG-AALD--P-A-SL--E-I---TRI-DP-YHPFRF
E05                N-I--IPGETI-R--H--HI-APFQD-SIVE--K-E-SL--E-F---TKI-KL-DHPLRN
E06                N-I--IPGETI-R--H--HI-APFQD-SIVE--K-E-SL--E-F---TKI-KL-DHPLRN
E04                NKL--FVGETF-R--H--HI-TPFEG-SVVE--K-E-SL--E-F---TKI-KL-DHPLRT
E03                EKL--VVGQTH-R--Y--HI-QPYQG-LEVE--D-E-SL--E-F---TKI-KL-DHPLRN
E02                E-F--LVGSTH-R--H--HI-IPFEG-LNVS--K-E-SL--E-F---TGI-DL-NHPLRN
Z3                 RRL---TGNNF-H--V--YL-KP--D-RLVD--P-E-AF--G-V---HGI--A-DE-F--
Z18                RRY---TGNN--H--I--YI-KP--D-RP-D--P-D-AI--K-V---HGI--T-DE-M--
Z2                 GKI---V-NPMSH--L---V-RP--D-RPIS--P-Q-AM--A-I---HRI--T-EA-M--
Z16                G-I---ITEKF-S--T--LV-YP--G-YFI---P-E-RIK-K-L---TGI--T-NA-M--
Z17                GEI---I-DKF-E--A--FA-NP--H-RPL---S-A-TII-E-L---TGI--T-DD-M--
AHIS1_ORF20        G-I-SWV-AQF-S--H--NV-KH-DV-YRVT--P-K-AM--E-I---NGL-RI-AN-L--
AHIS2_ORF27        K-I-DWIHCK--E--Y--RV-KQ-AV-YRVT--Q-G-AM--K-V---NRL-SL-VN-FT-
Z31                QKL---KKHSM-D--F--LL-FQFFN-FYVF--P-D-FLASQGFD-FNKV--F-ERQLRE
Z33                QIL---RDNVI-K--Y--RP-CQLYS-VPLF--Q-R-FLVKNKFN-LNQV--F-ERALLA
Z32                ELL---RKSSI-R--Y--TI-LQIVN-INVS--P-K-FLIQQGFD-FNKQ--L-ERNLID
E00_DNApolIII      GEF--FHLKEL-N--E--KI-KP-HR-IHLAN-S-K-AL--T-I---NGY--T-KK--K-
Z6                 ---------------A-LID-PL--G-IT-DWSPLK-AI--------------------


Z10                G-AK--K---TLREVQKD---LL--KIIS-RS-----DI---LIG-HSLQ-NDL---K--
```

```
Z11               N-CP------TYKEAL-DV--FLSENLIN-KN-----SI---LIG-HGLE-NDL---N--
Z12               N-AI------TFKEAQKK---TA--DIL--EG-----RI---LVG-HALK-HDL---E--
Z29               NFSK--KRLTALKYSYQKL--KY---LVD-VG-----VI---FVG-HGLK-NDF---R--
Z30               KISS--KHLTTLKSTYLKL--RF---LID-IG-----VK---FVG-HGLQ-KDF---R--
Z27               DRSP--YNVVPLKVAYKKL--RL---LVN-AG-----CI---FVG-HGLQ-KDF---R--
Z13               EKST--KRLVRRNVVYRKV--WL---LMQ-LG-----CV---FVG-HGLN-NDF---K--
Z28               TTSE--KYLTTHKRLLLRM--HV---LIQ-RG-----VT---FVG-HALH-NDF---T--
Z1                AKGE--NE--AAFAA-RIH--SL--FT-V-PK---T-CI---L-GYNNVRFDDEVT-RNI
Z15               EKGL--SE--PEFAA-NIL--AE--FS-Q-PN---T-CV---M-GYNNIRYDDEMT-RYT
Z5                TMGD--RE--AELATLEFL--KQ--WV-P-AG---K-SP---ICG-NSIG-QD----R-R
Z24               KLTE--RA--AELQTLDFL--KK--WV-P-KG---A-SP---ICG-NSIA-QD----K-R
Z26               TITL--QQ--AEYEFLSFV--RQ--QT-P-PG---L-CP---LAG-NSVH-ED----K-K
Z25               TVDL--AT--AEAMVLDYI--NE--HVKQ-PK---T-AP---LAG-NSIA-TD----R-A
Z9                EKTL--AQ--VEDELLEYI--QR--YI-PDKN---V-GV---LAG-NSVH-MD----R-L
E11               AIDE--KE--ALTSFFNDI--ND--AV-KSEYCS-R-AV---LVG-HNAH-FDL---G--
E12               ARPE--RE--VMQTMFGMV--RD--AI-KANHCK-R-AI---LVG-HNAH-FDL---G--
P11               PEEE--GE--ALRDIFRPI--RH--EI-SATGCT-R-SV---MVA-HNAH-FDL---G--
E10               PEEE--GE--ALREIFQPI--RR--EV-SITGCT-R-AV---MVA-HNAH-FDL---G--
P05               AVPE--AI--AFGEIFSAI--RK--EI-REQGCN-R-AI---LVG-HNAH-FDA---G--
P06               ADPE--EV--VFTDLLNAI--RK--QV-KVHDCT-R-AV---LVG-HNAH-FDH---G--
Z22               AVQE--EA--ALTEIFRGI--RK--AL-KANGCK-R-AI---LVG-HNSS-FDL---G--
P13               AVQE--EA--ALTEIFRGI--RK--AL-KANGCK-R-AI---LVG-HNSS-FDL---G--
P12               AVQE--EQ--ALTEIFRGI--RK--AL-KANGCK-R-AI---LVG-HNSS-FDL---G--
P14               AVSE--ES--AMTDIFRGV--RK--AL-KANGCK-R-AI---LVG-HNSS-FDL---G--
P15               AVSE--ET--ALTDIFRGV--RK--AL-KANGCK-R-AI---LVG-HNSS-FDL---G--
E01               AVSE--RE--ALTTLFRAV--RH--AV-RDSDCN-R-AV---LVG-HNAH-FDH---G--
P08               AKEE--KA--ALDHIFTPV--RA--AM-KKYGCQ-R-AI---LVG-HNAH-FDL---G--
P09               AKEE--KA--ALDHIFTPV--RA--AM-KKYGCQ-R-AI---LVG-HNAH-FDL---G--
Z21               AKEE--KL--ALEHIFTPV--RA--TM-KKYKCQ-R-AI---LVG-HNAH-FDL---N--
E09               ALDE--KQ--ALPKIFKPV--RN--AI-KNQGCN-K-AI---LVG-HNAH-FDL---K--
E07               ALDE--KQ--ALPKIFKPV--RN--AI-KNQGCN-K-AI---LVG-HNAH-FDL---K--
E08               ALDE--KQ--ALPKIFKPV--RN--AI-KNQGCN-K-AI---LVG-HNAH-FDL---K--
P10               AHPE--KD--ALATVFRGV--RE--EM-KLNACK-R-AI---LVG-HNAG-FDL---S--
P01               AVPE--RD--ALGKMFNPV--RQ--EL-RNTGCN-R-AV---LVG-HNAF-FDL---A--
P07               AVSE--RE--ALDKLLKPI--RS--AV-KNHGCT-R-AI---LVG-HNSF-FDL---N--
Z4                AVSE--YE--ALHEIFKVV--RK--GI-KASGCN-R-AI---MVA-HNAN-FDH---S--
Z20               AISE--KI--AIQSILKKV--RN--GI-KIQGCS-R-GI---VVA-HNAN-FDH---N--
Z19               AVSE--DH--ALKEIYKLV--RK--EQ-KAADCS-R-AI---IVA-HNAH-FDH---S--
Z23               AVSEL-D---AITGLFQMV--RR--GQ-KDADCQ-R-SI---IVA-HNAT-FDQ---S--
P04_DNApolIII     AVPE--DI--AITEMFKMV--RK--AV-KDNGCQ-R-AV---IVA-HNAT-FDQ---A--
A                 AVDET-E---AMKGLSKAI--SK--LQ-KAANCQ-R-SV---IVA-HNAS-FDM---S--
P02               AIAEDEKT--ALRRIFKGL--KE--VR-RNEDCR-Q-CV---LIG-HNAH-FDL---A--
P03               AIAEDEKT--ALRRIFKGL--KE--VR-RIEDCR-Q-CV---LIG-HNAH-FDL---A--
P00               AVDE--KK--ALQDLFAFV--EK--AV-NVAQCR-R-AV---LVG-HNAH-FDL---T--
E05               AINE--IE--ALDDLFKII--NK--TK-NKYECS-R-AI---LVG-HNAH-FDK---S--
E06               AINE--IE--ALDDLFKII--NK--TK-NKYECS-R-AI---LVG-HNAH-FDK---S--
E04               SLSE--VD--AIKDLFKII--NK--TK-NKYECS-R-AI---LVG-HNAH-FDS---S--
E03               AVEE---EV-ALKELFKII--NQ--TK-NKYECS-R-AI---LVG-HNAH-FDS---S--
E02               AVDE--KE--ALKDLFSII--NK--HK-NKYECS-R-AI---LVG-HNAH-FDL---G--
Z3                -LLD--KP--TFAEVADEF--MD--YI-R--G---A-E----LVI-HNAA-FDI---G--
Z18               -LAD--KP--EFKEVAQDF--LD--YI-N--G---A-E----LLI-HNAP-FDV---G--
Z2                -VAD--KP--WIEDVIPHY------Y-----G---S-EW---YVA-HNAS-FD-------
Z16               -LVG--QP--TIEEVLPEF--LE--FV----G---D-NI---VVG-HFVE-QDI---K--
Z17               -LQD--AP--DVVDVIRDF--RE--WI----G---D-DI---LVA-HNAS-FDM---G--
AHIS1_ORF20       --EN--AR--TLSQIWRQI--S------P--NIN-RNP---NIVG-HNVY-FDI---D--
AHIS2_ORF27       -QES--TL--TLKQIQGEI--KK--FY-PG-NYNGR-PFYHNLLG-HNVD-FDL---G--
Z31               QYDE--KQ--VVEK-IEDL-LQS--EI-ANSG---K-LV----IG-HNML-LDV---M--
Z33               GTAA--SE--PMQK-IP-LKMKP--NV-HMTG---K-LV----VG-HNSL-LDA---M--
Z32               KVNE--RH--STSKYLN-I-TTS--NI-LKKK---K-SV----VC-HNGM-ADL---V--
E00_DNApolIII     -WEN--AK--DMSKVLPKI--KE--MI-K-ES-----DL---LLG-QNLI-FDL---R--
Z6                ------------------------------LRDP-S-IT---KFL-HAGS-EDLE--V--
                                                                :    *

Z10               VMKLKHPL-VV--DTA--I----IYHH-K---AGDP-FK--PSL----K---YLS-----
Z11               VMRLFHNK-VI--DTA--I----LYS--R---T-K--FK--VSL----K---NLAF----
Z12               ALMLSHPKSLLR-DTSRHLPFRKLYA--K---G-K---T--PSL----K---KLTR----
Z29               VINIYVPSEQII-DTVH------LFHM-P---HHR---M--VSL----R---FLAW----
Z30               VVNLMVPKDQVL-DTVY------LFHM-P---RKR---M--ISL----R---FLAW----
Z27               IINLLVPPEQVV-DTVD------LFFL-SS--RQR---K--LSL----K---FLAW----
Z13               HININVPRNQIR-DTA-------IYFL-Q---GKR---Y--LSL----R---YLAY----
Z28               VLNVHVAESQII-DTVT------LMRL-G---TQR---M--LSL----Q---FLVK----
Z1                FYRNFYDPYAWSWQHD-------NSRW-D---LLDVMRAC-YAL----RPE-GINWPEND
Z15               FYRNFIDPYEYSWKNG-------NSRW-D---LLDLVRAC-YAL----RPE-GINWAYDD
Z5                FLFKYM-P-ELE-AYF-------HYR------YLDV--S---TL----K-ELARRW-K--
Z24               FLVKYM-P-DLA-DYF-------HYR------HLDV--S---TL----K-ELAARW-K--
Z26               FLDKYM-P-QFM-KHL-------HYR------IIDV--S---TV----K-ELCRRW-Y--
Z25               FIARDM-P-TLD-SFL-------HYR------MIDV--S---SI----K-ELCRRW-Y--
Z9                FMVREF-P-KVI-DHL-------FYR------IVDV--S---SI----M-EVARRH-N--
E11               FLNTAIKRNNIK-KNP-------FHGF-T---CFD---T--ATL----S---GLAY-G--
E12               FINASSERHNLK-RNP-------FHPF-S---CFD---T--ATL----S---GLAY-G--
P11               FVNAAVNRNQIK-RNP-------FHPF-S---CFD---T--STL----A---GLAY-G--
E10               FINAAVNRANIK-RNP-------FHPF-S---CFD---T--SGL----A---GLAY-G--
P05               FINAASMRADIK-RNP-------FHPF-S---HFD---T--ATL----A---GLAY-G--
P06               FVMAAVERCQIK-RNP-------FHPF-S---VFD---T--ATL----A---GLAF-G--
Z22               FLNAAVARTGIK-RNP-------FHPF-S---SFD---T--ATL----A---GLAY-G--
P13               FLNAAVARTGIK-RNP-------FHPF-S---SFD---T--ATL----A---GLAY-G--
P12               FLNAAVARTGIK-RNP-------FHPF-S---SFD---T--ATL----A---GLAY-G--
P14               FLNAAVARNDLK-RNP-------FHPF-S---SFD---T--ATL----A---GLAY-G--
P15               FLNAAVARLDMK-RNP-------FHPF-S---SFD---T--ATL----A---GLAY-G--
E01               FMQAAAARQDLK-RNP-------FHPF-S---SFD---T--VTL----A---GLAY-G--
P08               FVNAAVARTGHK-RNP-------FHPF-S---VFD---T--VTL----A---GIAY-G--
P09               FVNAAVARTGHK-RNP-------FHPF-S---VFD---T--VTL----A---GIAY-G--
Z21               FLNAAVTRTAYK-RNP-------FHQF-S---VFD---T--VTL----A---GMAY-G--
E09               FINAAAARSGIK-RNP-------FHPF-S---TFD---T--VSL----A---GLAY-G--
E07               FINAAAARSGIK-RNP-------FHPF-S---TFD---T--VSL----A---GLAY-G--
E08               FINAAAARSGIK-RNP-------FHPF-S---TFD---T--VSL----A---GLAY-G--
P10               FIHQGAERAGLK-RNP-------FHPF-S---SLD---T--VTL----S---AMAF-G--
P01               FLNAAVTRTGFK-RNP-------FHPF-S---SFD---T--ATL----G---GLAY-G--
P07               FLNAAIDRCTYK-RNP-------FHAF-S---SFD---T--VTI----A---GLAY-G--
```

```
Z4                FMMAAAERASLK-RNP-------FHPF-A---TFD---T--AAL----A---GLAL-G--
Z20               FLMAAIQRVKIK-NNP-------FHPF-A---TFD---T--AAL----S---GLVV-G--
Z19               FVMAASERCKLK-RVP-------FHPF-A---TFD---T--ATL----S---GLAF-G--
Z23               FVMAAAERTGVK-RNP-------FHPF-G---MFD---T--ASL----A---GLMF-G--
P04_DNApolIII     FLQAAVKRINTK-RDP-------FHPF-A---MFD---T--ATL----A---GFMY-G--
A                 FVNAALARTNVK-RSP-------FHPF-V---SFD---T--TSL----A---ALTL-G--
P02               FLNAAVLRTNNKSHNP-------FHNF-S---VFD---T--VTL----S---ALAF-G--
P03               FLNAAVLRTNNKSHNP-------FHNF-S---VFD---T--VTL----S---ALAF-G--
P00               FIQAAMKRCKLK-KSP-------FHAF-T---CFD---T--ATL----A---AIVF-G--
E05               FLDAAVERNNIK-KTP-------FHKF-S---VID---T--VSL----G---ALAT-G--
E06               FLDAAVERNNIK-KTP-------FHKF-S---VID---T--VSL----G---ALAT-G--
E04               FLTAAYERNKIK-KTP-------FHKF-S---VID---T--VSL----G---VLAT-G--
E03               FLNAAIERNGIK-RSP-------FHPF-S---VLD---T--VTL----G---ALAT-N--
E02               FLNESIKRNNIK-RSP-------FHPF-S---VFD---T--VSI----G---GIFT-G--
Z3                FMDYEF-SL-LK-RDI-------PKT--N---TFCKV-TDSLAV----A---RKMFPG--
Z18               FMDYEF-RK-LN-LNV--------KT--D---DICLV-TDTLQM----A---RQMYPG--
Z2                ----R--RV-LP--EM--------PG--E---WIC---T--MKL----A---RRLWPGI-
Z16               FIN-KYTKQ-YR-GKK-------FRN--P---SLC---T--LKL----A---RKVFPGL-
Z17               FLNVAYKKL-LE-VEK-------AKN--P---VID---T--LEL----G---RFLYPEF-
AHIS1_ORF20       FLHGIPFESHKKPRNV-------FDTY-A---IAN---V--LKA----A---RIPY-K--
AHIS2_ORF27       FTQGYPFESGKRPQNI-------YDTM-R---IAQ---A--LKA----A---GLRY-G--
Z31               HTVHQF-YCPFK-EMT-------TCVF-P--RLLD---T--KLM----A--STQPF-KP-
Z33               YMYHYF-FSHFK-DKF-------NALF-P--RIMD---T--KLL----A--QALRFEKSD
Z32               YLFSLF-EGKFS-ELC-------LSSF-K--SIYD---T--KLLY-LKS--DDLPSNSSS
E00_DNApolIII     FLQKAS-LTEID-FPP-------YIDT-K---AMA---D---FL----V---RKNWLR--
Z6                FLN-VF---GEL-PQP-------LIDT-Q--ILAAFCGRPMSWGFASMVEEYSGVTLDKS


Z10               -----ETFLNKS---------IQNG-E----HDSVEDA-RACLELTKLKIL
Z11               -----EV-LSRK---------IQNG-E----HDSSQDA-IATMDVVKVKIG
Z12               -----EV-LKIS---------IQEG-E----HSSVEDA-RATMLLYKKEKT
Z29               -----HF-LGTK---------IQ-S-ET---HDSIEDA-RTTLQLYKHYLK
Z30               -----YF-LDLK---------IQ-G-ET---HDSIEDA-RTALQLYRKYLE
Z27               -----YL-LDEE---------IQLT-E----HDSIEDA-LTALKLYDCYDK
Z13               -----VL-LGMN---------IQEG-N----HDSIEDA-HTALILYKKYLH
Z28               -----EI-LGET---------IQ-M-E---AHDSVVDA-RYALKLYRKYLE
Z1                DGLP-SFRLEHLT-KA---NGIEHS-N---AHDAMADV-YATIAMAKLVKT
Z15               DGMP-SFRLEKLT-KA---NGIEHE-N---AHDAMADV-YATIAMAKLIKE
Z5                ---P-EI-LDGF----------TKQ-G---THQAMDDIRESVAELAYYREH
Z24               ---P-EI-LEGF----------KKE-N---THLALDDIRESIKELAYYREH
Z26               ---P-EE-YE-FA--P------KKA-A---SHRALDDISESIKELQFYRNN
Z25               ---P-RI-Y--FG-QP------PKG-L---THRALADIHESIRELRFYRRT
Z9                ---P-AL-QARN---P------KKE-A---AHTAYSDIKESIAQLQWYMDN
E11               ---Q-TV-LAKAC-KS---AQIEFN-SDK-AHSAIYDA-EKTAELFCAIVN
E12               ---Q-TV-LARAC-ET---AEIDFN-NDE-AHSARYDA-QKTAELYCKIVN
P11               ---Q-TV-LAKAC-QA---AKMDFD-NRS-AHSALYDA-VKTAELFCGIVN
E10               ---Q-TV-LAKAC-KA---AGIDFN-NND-AHSALYDT-VKTAELFCGIVN
P05               ---Q-TV-LAKAC-RT---AGIDFD-NNE-AHSAAYDA-AKTAELFCNIVN
P06               ---Q-TV-LAKAC-QT---AGISFN-NAE-AHSAAYDT-GKTAELFCLIVN
Z22               ---Q-TV-LAKAC-QA---AGMEFD-NRE-AHSARYDT-EKTAELFCGIVN
P13               ---Q-TV-LAKAC-QA---AGMEFD-NRE-AHSARYDT-EKTAELFCGIVN
P12               ---Q-TV-LAKAC-QA---AGMEFD-NRE-AHSARYDT-EKTAELFCGIVN
P14               ---Q-TV-LARAC-QS---ADIDFD-GRE-AHSARYDT-EKTAELFCGIVN
P15               ---Q-TV-LAKAC-QA---ADIDFD-GRE-AHSARYDT-EKTAELFCGIVN
E01               ---Q-TV-LARAC-QA---AGITFD-SRA-AHSALYDA-EKTAELFCAIVN
P08               ---Q-TV-LARAA-TA---AGLGWD-ANE-AHSAVYDT-EQTARLFCTIAN
P09               ---Q-TV-LARAA-NA---AGLGWD-ANE-AHSAVYDT-EQTARLFCTIAN
Z21               ---Q-TV-LARAV-QA---AGLDWN-AAD-AHSAVYDA-EKTAHLFCTITN
E09               ---Q-TV-LAHSI-RS---AGLDWD-SNE-AHSAVYDA-EMTAMLFCKIVN
E07               ---Q-TV-LARSI-RS---AGLDWD-SNE-AHSAVYDA-EMTAMLFCKIVN
E08               ---Q-TV-LARSI-RS---AGLDWD-SNE-AHSAVYDA-EMTAMLFCKIVN
P10               ---Q-TV-LARAV-QA---AGMEWS-NSE-AHSAVYDA-EKTADLFCHVIN
P01               ---Q-TV-LAKAV-KA---AGIEWD-AKE-AHSAIYDA-EKTAELFCAIVN
P07               ---Q-TV-LARAC-EA---AGLGWN-NSE-AHSAVYDA-EQTARLFCKVIN
Z4                ---Q-TV-LSKAC-QT---AGMDFD-STQ-AHSALYDT-ERTAVLFCEIVN
Z20               ---Q-TV-LSKAC-KA---IGLSFD-NHQ-AHSALYDT-LQTANLFCELVN
Z19               ---Q-TV-LAKAC-KT---AGMEFD-NRE-AHSALYDT-QKTAELFCTIVN
Z23               ---Q-TV-LVKAC-QA---AKIPFD-GKQ-AHSALYDT-ERTAKLFCYMVN
P04_DNApolIII     ---Q-TV-LVKAC-QM---AQIPFD-GKQ-AHSALYDT-ERTAELFCAMVN
A                 ---Q-TV-LVKAC-EA---ANIAFD-QKQ-AHGALYDA-QRTAELFCFMMN
P02               ---Q-TV-LARAC-KA---AGLDFD-GND-AHSALYDT-QKTAELFCYILN
P03               ---Q-TV-LARAC-KA---AGLDFD-GND-AHSALYDT-QKTAELFCYILN
P00               ---K-TV-LAKAL-RA---ANLPFD-KKE-AHSAIYDA-KCAAELFCKITN
E05               ---Q-TV-LARIC-DE---LSIEYD-NNE-AHSAAYDT-KVTAEVFCKIVN
E06               ---Q-TV-LARIC-DE---LSIEYD-NNE-AHSAAYDT-KVTAEVFCKIVN
E04               ---Q-TV-LARIC-DE---LDIDYD-NDE-AHSAAYDS-DVTAKVFCSIIN
E03               ---Q-TV-LARIC-EL---LDIDYD-SKE-AHSAAYDS-DVTAQVFCKIIN
E02               ---Q-TV-LARVA-RE---LGFDYE-N-ELAHSASYDS-EITAK-------
Z3                --KR-NS-LDALC-AR---YEIDNS-KRTL-HGALLDA-QILAEVYLAMTG
Z18               --KR-NN-LDALC-DR---LGIDNS-KRTL-HGALLDA-EILADVYLMMTG
Z2                --KY-SN-M-ALYKTR---LNVQTP-PGLHHHRALYDC-YITAALLIDIMN
Z16               --KK-YS-LKEIA-EN---FGFE-T-NG--VHRALKDA-TLTAEIFIKILE
Z17               --KN-HR-LNTLC-KK---FDIELT-QH---HRAIYDT-EATAYLLLKMLK
AHIS1_ORF20       ---K-LN-LAAVC-RY---HNIDFD-ESE-AHEALYDA-IKSVEVLHAQLS
AHIS2_ORF27       ---K-VK-LQSCC-DF---HGITYN-EDL-AHGALYDA-ERTVEVFQAQLK
Z31               --FN-PPKVES-A-EGF--PSYDTA-SEQL-HEAGYDA-YITGLCFISMAN
Z33               --KTVPPELRGFI-EPWMNP-LEDE-SENVYHNAGFDS-YVTGEVFLKLAH
Z32               --QRSNVSLNSLI-EC---P-YKSMMSRKRPHEAGKDS-YDTALLFVYYVM
E00_DNApolIII     ---R-SK-MDYLV-EH---YNVEVK-GR--AHTALVDC-YRTYEVFKKLLS
Z6                -ESRTDW-LARP-------LTE-----R-QCEYAAADVWYLLPITAKLMVE
                                              . :  *
```

Table A.4: Predicted A-HIS1 RBS. A-HIS1 RBS patterns and positions found using a RBS region length of 50. The position is that of the first nucleotide of the RBS pattern.

ORF	RBS (AGGAGG)		RBS (GGAGAA)	
	Pattern	Position	Pattern	Position
4	AGGAAG	2532	-	-
5	CGGAGA	2737	GGAGAA	2736
9	AGGAAA	3924	-	-
13	AGGAGA	5573	GGAGAA	5572
15	AGGAAA	8828	-	-
16	AGGAAA	10712	-	-
19	CGGAGA	12405	GGAGAA	12404
20	ACGAGG	12978	-	-
22	CGGAGG	13314	GGAGGA	13313
23	AGGAAA	13566	-	-
24	-	-	GGAGCA	14208
25	AGGAGA	14381	GGAGAA	14420
28	AGGAGA	16555	GGAGAA	16554
30	AGGAGA	18073	GGAGAA	18072
37	ACGGGG	21529	-	-
40	ACGAGG	24414	-	-
42	AGGAGA	25074	GGAGAA	25075
47	-	-	GGAGAA	26645
48	AGGAGA	27270	GGAGAT	27271
50	AGGAGA	28074	GGAGAT	28075
51	CGGAGG	28459	GGAGGC	28460
53	ACGGGA	29665	-	-
57	-	-	GGAGTA	31861
59	AGGAGA	32832	GGAGAA	32833
61	-	-	GGAGAT	33470
62	AGGAGA	34168	GGAGAA	34169
64	AGGCGA	35752	GGCGAT	35753
65	AGGAAA	36408	-	-
66	AGGAGA	36582	GGAGAA	36583
67	GGGAGC	37226	GGAGCG	37227
70	AGGAAA	37855	-	-
73	AGGAAA	39096	-	-
78	-	-	GGAGCA	42975
81	AGGAAA	45550	-	-
82	-	-	GGAGAA	45822
83	AGGAGA	46544	GGAGAG	46545
84	AGGAGA	47042	GGAGAA	47041
85	-	-	GGCGAA	47232
86	CGGAGA	47679	GGAGAA	47678
87	AGGAGA	47783	GGAGAT	47784
90	AGGAGA	50298	GGAGAT	50297
93	AGGAGC	51561	GGAGCA	51560
94	AGGAGC	52700	GGAGCC	52699

Table A.5: Predicted A-HIS2 RBS. A-HIS2 RBS patterns and positions found using a RBS region length of 50.

ORF	RBS (AGGAGG)		RBS (GGAGAA)	
	Pattern	Position	Pattern	Position
2	ATGGGG	1663	GGGGAT	1661
3	CGGAGG	1892	GGAGGA	1891
4	CAGAGG	2157	-	-
5	ACGAGG	2442	-	-
6	ATGAGG	2831	-	-
8	CCGAGG	3347	-	-
10	AGGAGT	3848	-	-
15	CCGGGG	5288	-	-
16	AGGAGA	5671	GGAGAT	5670
17	AAGGGG	6035	GGGGAC	6033
18	AGGGTG	6123	-	-
22	CAGAGG	11843	-	-
24	AAGAGG	12402	-	-
26	CGGAGG	13395	GGAGGA	13394
27	AGGTGG	13982	-	-
29	AGGGGG	14501	GGGGGT	14500
31	AAGAGG	15591	-	-
32	AGGGTG	16362	-	-
34	ATGATG	17902	-	-
38	-	-	GGGGAT	21461
39	CTGGGG	21719	-	21703
40	AGGTGG	21617	GGTGAT	21640
41	AAGAGG	21965	GGGGAT	21993
42	CGGAGG	22274	GGAGGA	22273
44	AAGAGG	22470	GGAGTA	22459
45	AGGGGC	22720	-	-
46	-	-	GGGGAT	23081
48	-	-	GGGGTA	23407
50	-	-	GGGGAT	23802
51	ACGAGA	23961	-	-
54	CGGAGG	24592	GGAGGG	24593
55	AGGATG	24869	-	-
56	CGGAGG	25512	GGAGGA	25513
57	ATGAGG	25710	-	-
58	CAGAGG	27843	GGTGAA	27834
59	AGGTGG	27970	GGTGGA	27971
60	ACGAGG	28122	-	-
61	ACGAGG	28712	-	-
62	-	-	GGGGAT	29090
63	ATGAGG	29248	-	-
64	CAGAGG	29463	GGAGCA	29430
65	-	-	GGAGCA	29820
66	AGGTGG	30219	-	-
68	AGGGGG	31437	GGGGGA	31438
71	AGGAGA	33286	GGAGAT	33287
73	AGGATG	33961	-	-
74	AGGGTG	35312	-	-
75	AAGAGT	36058	-	-
77	AGGAGA	36185	GGAGAA	36186
79	GTGAGG	37255	-	-
80	AAGAGG	37719	-	-
83	GCGAGG	39225	-	-
85	ATGAGG	41667	-	-
86	CGGAGG	42265	GGAGGA	42266
87	ATGAGG	45607	-	-
89	AAGGTG	45788	GGTGAA	45790
90	ACGAGG	46087	-	-
91	AAGAGG	46978	-	-
95	AGGAGT	47809	-	-
96	-	-	GGGGAA	48063
97	GTGAGG	49491	-	-
98	CAGAGG	49570	GGGGCA	49548
100	AGGAGA	50826	GGAGAA	50825
102	AAGAGG	52491	GGAGAA	52456
103	GTGAGG	56778	-	-

Table A.6: A-HIS1/2 transcriptional terminator data. Region designations as provided with software; F = intragenic, between two + strand genes, G = intergenic, at least 50 bases from an end of a gene.

Terminator detected	Start	End	Length	Strand	Confidence	Hairpin score	Tail score	Region	Upstream ORF	ORF-terminator distance (bases)	Comment	Status
A-HIS1												
1Term1	4469	4479	11	+	76	-6.8	-4.56702	F	N/A	-	no ORF in vicinity	False
1Term2	8631	8647	17	+	78	-8.7	-3.74178	F	N/A	-	no ORF in vicinity	False
1Term3	29300	29327	28	+	91	-7.6	-4.77719	F	51	26		Putative
1Term4	34348	34371	24	+	93	-8.8	-4.54449	F	62	24		Putative
1Term5	37512	37539	28	+	90	-11.2	-4.05607	F, G	67	11	in ORF 68	False
1Term6	40211	40236	26	+	86	-6.7	-5.07523	F	73	22	upstream ORF maj head	Putative
1Term7	42340	42365	26	+	81	-7.5	-4.59993	F	76	24		Putative
1Term8	43124	43151	28	+	100	-11.2	-5.14163	F	78	7		Putative
1Term9	47998	48025	28	+	100	-12.6	-4.37198	F	87	12		Putative
1Term10	49999	50026	28	+	100	-12.6	-3.70811	F	89	23	upstream ORF VSP	Putative
A-HIS2												
2Term1	6763	6780	18	+	100	-10.5	-5.10933	F	19	9		Putative
2Term2	24481	24500	20	+	100	-11.4	-4.77165	F	52	22		Putative
2Term3	25714	25739	26	+	100	-11.3	-5.18726	F	56	18		Putative
2Term4	31121	31140	20	+	100	-11.4	-4.95302	F	66	30		Putative
2Term5	33923	33945	23	+	100	-13.4	-3.6215	F	72	89		Putative
2Term6	39180	39206	27	+	78	-9.3	-3.66883	F	82	34	upstream ORF maj head	Putative
2Term7	41614	41638	25	+	100	-14.3	-4.68593	F	84	166		Putative
2Term8	42417	42447	31	+	100	-12.7	-3.77722	F	86	7	overlap with 2Term8	Putative*
2Term9	42423	42440	18	+	100	-12.7	-4.80656	F	86	13		Putative
2Term10	48267	48296	30	+	76	-10	-3.27808	F	96	8	overlap with 2Term11	Putative*
2Term11	48269	48294	26	+	93	-12.6	-3.19859	F	96	10		Putative
2Term12	49771	49806	36	+	100	-17	-3.91798	F	98	0		Putative

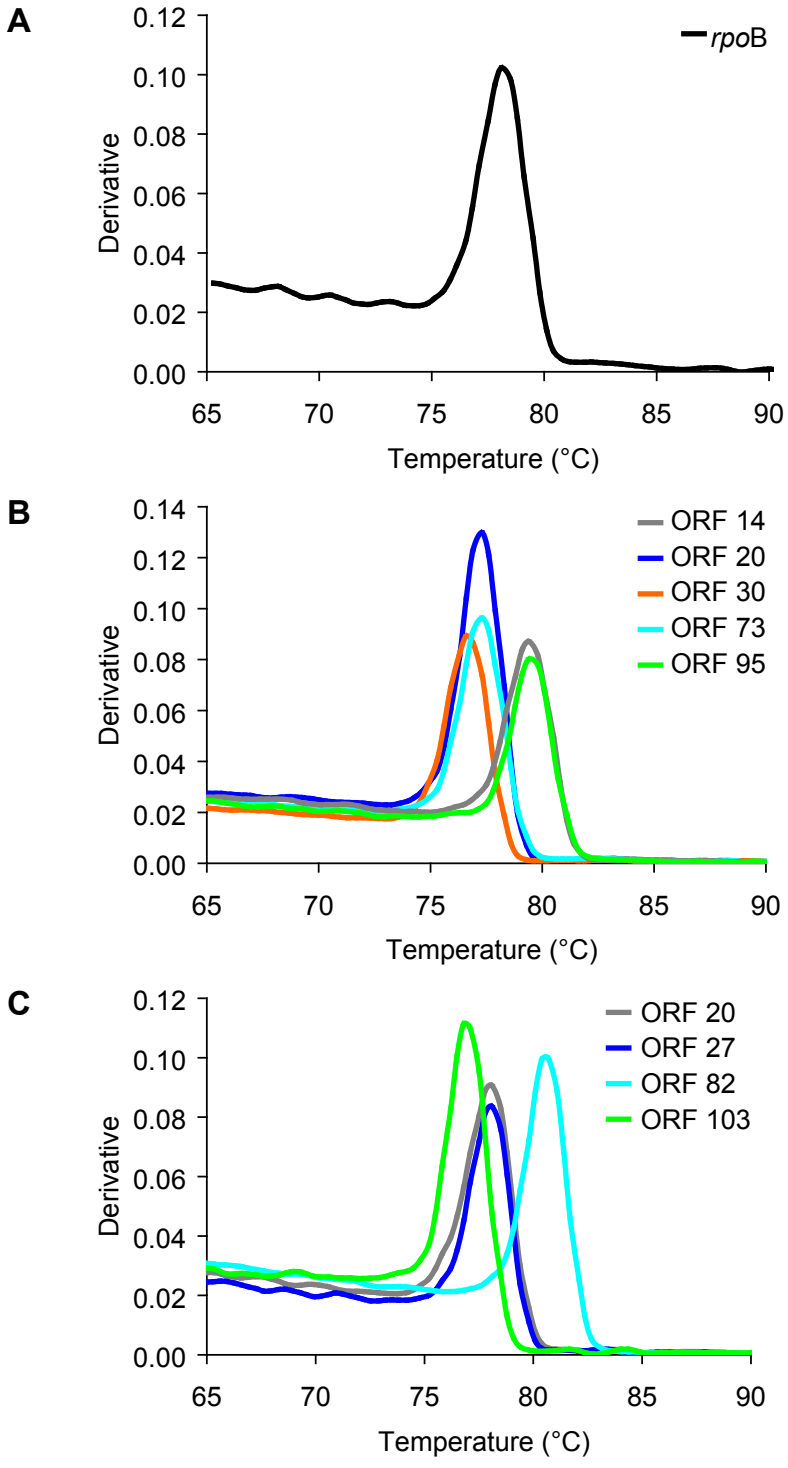


Fig. A.4: qPCR melting curves. Melting curves for (A) *A. marina rpo*B and genes from (B) phage A-HIS1 and (C) phage A-HIS2: ORFs 14 and 20, mtDNA polymerase); 20 and 27, RNase T, 73 and 82 (major capsid), 95 and 103 (putative cell surface protein), respectively. Phage A-HIS1 ORF 30 (DNA polymerase I-like FEN) was included in intra-plate calibration (Fig. 5.9).

APPENDIX B

Publications

Pigment composition and adaptation in free-living and symbiotic strains of *Acaryochloris marina*

Yi-Wah Chan[1], Anja Nenninger[2], Samuel J.H. Clokie[3], Nicholas H. Mann[2], David J. Scanlan[2], Anna L. Whitworth[4] & Martha R.J. Clokie[5]

[1]MOAC Doctoral Training Centre, University of Warwick, Coventry, UK; [2]Department of Biological Sciences, University of Warwick, Coventry, UK; [3]School of Biomedical and Clinical Laboratory Sciences, University of Edinburgh, UK; [4]Chemistry Department, University of Warwick, Coventry, UK; and [5]Department of Infection, Immunity and Inflammation, Medical Sciences Building, University of Leicester, Leicester, UK

Correspondence: Martha R.J. Clokie, Department of Infection, Immunity and Inflammation, Medical Sciences Building, University of Leicester, PO Box 138, Leicester LE1 9HN, UK. Tel.: +44 11625252959; fax: +44 1162525030; e-mail: mrjc1@le.ac.uk

Received 19 September 2006; revised 14 February 2007; accepted 26 February 2007. First published online 28 April 2007.

DOI:10.1111/j.1574-6941.2007.00320.x

Editor: Patricia Sobecky

Keywords
MBIC11017; CCMEE5410; light intensity; adaptation; phycobiliproteins; chlorophyll *d*.

Abstract

Acaryochloris marina strains have been isolated from several varied locations and habitats worldwide demonstrating a diverse and dynamic ecology. In this study, the whole cell photophysiologies of strain MBIC11017, originally isolated from a colonial ascidian, and the free-living epilithic strain CCMEE5410 are analyzed by absorbance and fluorescence spectroscopy, laser scanning confocal microscopy, sodium dodecyl sulfate polyacrylamide gel electrophoresis and subsequent protein analysis. We demonstrate pigment adaptation in MBIC11017 and CCMEE5410 under different light regimes. We show that the higher the incident growth light intensity for both strains, the greater the decrease in their chlorophyll *d* content. However, the strain MBIC11017 loses its phycobiliproteins relative to its chlorophyll *d* content when grown at light intensities of 40 $\mu E\,m^{-2}\,s^{-1}$ without shaking and 100 $\mu E\,m^{-2}\,s^{-1}$ with shaking. We also conclude that phycobiliproteins are absent in the free-living strain CCMEE5410.

Introduction

Miyashita *et al.* (1996) first discovered *Acaryochloris marina* (hereby referred to as *A. marina*) strain MBIC11017 on the coast of the Palau Islands in the western Pacific. *Acaryochloris marina* is distinguished from other cyanobacteria by virtue of possessing mostly Chl *d* as opposed to Chl *a*. New strains of *A. marina* have subsequently been isolated in a variety of habitats and locations, including the Awaji strain, which grows as an epiphyte on the thalli of red algae from the Awaji Island seashore, Japan (Murakami *et al.*, 2004; Akimoto *et al.*, 2006). A recent paper by Ohkubo *et al.* (2006) reveals that *Acaryochloris* sp. are in fact widespread as epiphytes on not only red, but also on brown and green macroalgae. In 2005, an endolithic 'free-living' strain CCMEE5410 was discovered in the eutrophic, hypersaline conditions of the Salton Sea, USA (Miller *et al.*, 2005). To add to the habitats of *A. marina*, a putative endolithic strain has been identified by de los Ríos *et al.* (2007) growing in a biofilm loosely adhered to fissures in granite rocks in continental Antarctica. Identification was made on the basis of both morphology and 16S rRNA gene, where the strain clustered with *Acaryochloris* in a phylogenetic tree.

Although the subject of the habitat of *A. marina* has been of some interest (Kühl *et al.*, 2005; Larkum & Kühl, 2005), the majority of research on this organism has been focused on understanding its mechanisms of photosynthesis (Hu *et al.*, 1998; Mimuro *et al.*, 2004; Chen *et al.*, 2005b), the effects of iron on growth and physiology (Chen *et al.*, 2005a; Swingley *et al.*, 2005), photosystem properties (Kumazaki *et al.*, 2002) and pigment composition (Akiyama *et al.*, 2001; Kobayashi *et al.*, 2005). *Acaryochloris marina* uses a combination of phycobiliprotein (PBP) rods and photosystem complexes to harvest light. Initial studies of the PBPs indicated that the rods consist of three homo-hexameric units containing phycocyanin (PC) with a terminal hetero-hexameric unit containing both PC and allophycocyanin (APC) (Marquardt *et al.*, 1997; Hu *et al.*, 1999). Although the presence of PBPs has been verified, little is known about their response to different light regimes.

Details of *A. marina* growth conditions are often scant but strain MBIC11017 has been cultured under a range of conditions in several laboratories, at light intensities ranging from 5 to 80 $\mu E\,m^{-2}\,s^{-1}$ and either under continuous light or light/dark regimes. Temperatures have ranged from 20 to 30 °C. Aeration and, in particular, shaking and stirring of

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

cultures have been used sporadically. Strain CCMEE5410 has only been cultured under 15 and 50 $\mu E\,m^{-2}\,s^{-1}$ of continuous fluorescent light (Miller *et al.*, 2005).

Observations in this laboratory from initial culturing of strains MBIC11017 and CCMEE5410 indicated a significant difference in the pigmentation of these two strains of *A. marina*. Under a range of temperatures and light intensities, the CCMEE5410 cultures were yellow, becoming yellow-green as cultures aged. In contrast, the MBIC11017 cultures were bright lime green, turning a dark green the more dense the culture. However, under certain conditions, MBIC11017 cultures became yellow and virtually indistinguishable from the CCMEE5410 cultures.

The subject of algal and cyanobacterial pigments and species distribution in the water column has been debated for many years, starting with Engelmann putting forward his theory of chromatic adaptation in 1883, in which he stated that algal pigments are controlled by the quality of light at a particular depth (Engelmann, 1883). An alternative theory was proposed by Oltmanns (1892), in which he proposed that it was the quantity of light and not the quality which controlled the depth distribution.

This is the first study to focus on the effects of growth irradiance on *A. marina* pigment composition. The two strains of *A. marina* were isolated from very different habitats and light regimes. Their pigment composition and photosynthetic physiology were measured under a range of growth conditions and the observations were related to the habitat in which they are found.

Materials and methods

Culture conditions

Strains MBIC11017 and CCMEE5410 were cultured in autoclaved artificial sea water (ASW) (Wyman *et al.*, 1985) in 1 L polycarbonate NALGENE® 4105 Fernbach culture flasks with aeration, stirring and under 30–50 $\mu E\,m^{-2}\,s^{-1}$ of continuous white fluorescent light (Osram Ltd, L65/80W/23). BD Falcon polystyrene tissue culture flasks of 50 mL containing 44 mL of autoclaved ASW were inoculated with 6 mL of an exponentially growing culture and grown for a period of 14 days: at (1) 23 °C and (2) 28 °C under continuous white fluorescent light and (3) 28 °C with a 12 h/12 h light/dark regime at 10, 20, 30 and 40 $\mu E\,m^{-2}\,s^{-1}$ growth light intensity. For the experiments which involved shaking, MBIC11017 cultures were grown at 28 °C under 25, 50 and 100 $\mu E\,m^{-2}\,s^{-1}$ at 125 r.p.m. on a New Brunswick Scientific Innova™ 2000 platform shaker. Light intensity was determined using a Skye Instruments Ltd SKP 200 with lux sensor. The nomenclature for samples used in this paper denotes the strain and light intensity at which they were grown. For example, M10 refers to strain MBIC11017 grown at 10 $\mu E\,m^{-2}\,s^{-1}$.

Absorbance and fluorescence measurements

Absorbance and fluorescence spectra were recorded at room temperature on a UV 500 UV/VIS spectrometer using VISION32 software and a Spectronic Instruments 8100 Series spectrofluorometer, respectively. Absorbance data were normalized to cell light scattering. The cell light scattering data was generated from absorbance data from boiled MBIC11017 cells. Absorbance and fluorescence spectra were smoothed using the Savitzky–Golay algorithm (Savitzky & Golay, 1964). Samples grown under static conditions were shaken before sampling.

Protein extraction, MS and Western blotting

Cells were harvested and pelleted *c.* 300 h after inoculation, flash frozen in liquid nitrogen and stored at −20 °C. Thawed cells were washed and resuspended in 10 mM HEPES pH 7.4, bead beaten using 0.17–0.18 mm Glasperlen (B. Braun Biotech International GmbH) for 15 min at room temperature and the crude cell extract was removed after centrifugation at 16 000 ***g***. The protein concentration of each sample was assessed using a BCA Protein Assay Kit in 96-well plates (Becton Dickinson) and a Labsystems iEMS Reader MF with ASCENT Software Version 2.4.2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) samples were prepared after thawing by adding a 5× denaturing mix (5% (w/v) SDS, 27% (v/v) glycerol, 98 mM Tris-HCl pH 6.8, 1.3% (v/v) β-mercaptoethanol and 0.2% (w/v) bromophenol blue) and heating at 100 °C for 3–5 min. Polyacrylamide gels of 15% were cast using a BioRad Mini-Gel kit. Bilin-containing bands were observed under UV and imaged in GENEFLASH on a Syngene Bio Imaging transilluminator unit. Visualization of bilin-containing bands under UV was enhanced by incubating the gels in 2 M glycine with 0.2 M $ZnSO_4$ for 30 min (Berkelman & Lagarias, 1986). The gels were stained with Coomassie-blue in 3% (v/v) acetic acid and destained in water.

Liquid chromatography-MS/MS was used to determine which proteins were present in the bands observed on the polyacrylamide gels. Individual bands were excised and treated according to the protocol of Aitken & Learmonth (2002). Briefly, gel pieces were incubated with three changes of 200 mM ammonium bicarbonate (ABC)/50% (v/v) acetonitrile (ACN) at 30 °C for 30 min to remove SDS. The proteins were reduced by incubation with 20 mM DTT (dithiothreitol)/0.2 M ABC/50% ACN at 30 °C for 1 h. After several washes to remove DTT, the cysteine residues were alkylated by incubation with 50 mM iodoacetamide in fresh 200 mM ABC/50% ACN for 20 min in the dark. After further washing in 20 mM ABC/50% ACN, the band was

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

cut into small (1×2 mm) pieces and shrunk with the addition of 100% ACN. The gel pieces were then rehydrated with 0.5 μg modified sequencing grade trypsin (Promega), made up in 50 mM ABC (pH 7.8), on ice for 15 min. Then 100 μL of 50 mM ABC, pH 7.8, was added to cover the pieces and incubated overnight at 30 °C. The peptides contained within the supernatant were transferred to another tube and further extraction was achieved by the addition of another 100 μL 200 mM ABC/50% ACN and sonication for 30 min at 35 °C. The supernatants were collected, combined and lyophilized in a centrifugal evaporator for 30 min.

The aqueous solution containing digested peptides was passed through a 0.22 μm filter (Millipore, MA), then loaded onto a 15-cm, 75-μm internal diameter, C18 media column (LC packings, Dionex, CA) and eluted with a 5–95% gradient of water: acetonitrile (0.1% formic acid) over 60 min at a flow rate of 0.2 μL min^{-1} directly into a Thermofinnigan LCQ DECA quadropole ion trap instrument using online nanoHPLC-MS/MS. MS/MS data was collected using data-dependent acquisition experiments on the major ion species using the accompanying EXCALIBUR software (Thermo Electron Corp., San Jose, CA).

The raw data from the LCQ was processed using SEQUEST (Eng *et al.*, 1994) to establish parent: daughter ion relationships, producing a series of .dta files for each LC-MS/MS run. The merged .dta files were then submitted to the MASCOT server at www.matrixscience.com for searching against the entire NCBI nr database.

Samples for Western blotting were analyzed on 16% SDS-PAGE gels. To increase the resolution for small proteins, 1.25% (w/v) solid Bis-Tris was added and the concentration of ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) was doubled to 0.2% (w/v) and 0.005% (w/v) respectively. The gel was blotted overnight onto Hybond P (Amersham Biosciences) using the Towbin-buffer system. The membrane was blocked in PBS-T (PBS with 0.1% Tween) with 3% skimmed milk powder for 1 h followed by three 30-min washes in PBS-T. The primary antibody [rabbit anti-APC polyclonal antibody, unconjugated from AbD Serotec (Biogenesis) obtained from being challenged with APC and PC in *Spirulina platensis*] was incubated for 1 h at 1 : 1000 and the antirabbit IgG HRP conjugate antibody (Promega) was incubated at 1 : 10 000 for 1 h. This was washed with PBS-T and then the detection reagents were added for 1 min (Amersham, ECL-kit). The blots were developed using an AGFA Curix 60 on Fuji Medical X-ray film for 5–30 s.

Laser scanning confocal microscopy (LSCM)

Liquid cell culture samples were spotted onto 1% agar-ASW plates and left to dry on the bench for 10–20 min. LSCM was performed using a Leica SP2 and a Leica DM RE7 upright microscope. Blocks of agar with adsorbed cell samples were cut from the agar plate and placed into the well of a custom built sample holder. A coverslip was placed flush over the sample and then imaged using a $\times 63$ oil immersion lens. Phycobiliprotein and chlorophyll *d* fluorescence were detected between 640–670 and 700–750 nm respectively by excitation with a 633 nm red laser. The same gains were used for the PBP (400 ± 1 V) and Chl *d* (490 ± 1 V) channels to compare shaken MBIC11017 cultures grown at 25, 50 and 100 μE m^{-2} s^{-1}. The offset in both channels was set to $2 \pm 1\%$. Where comparison between images from the same channel was made, the same gains and offset conditions were used. In Volocity (version 3.7.0) from Improvision® the maximum intensity projection (MIP) contribution was adjusted for all images from 100% (acquired image) to 200%, to give maximum contrast.

Results and discussion

Absorbance spectra

Whole cell absorbance spectra for strain MBIC11017, grown at 28 °C, under 10–20 μE m^{-2} s^{-1} of continuous white fluorescent light were consistent with previously published spectra (Miyashita *et al.*, 1996, 1997; Mimuro *et al.*, 2000). Strain CCMEE5410 (Fig. 1d) exhibited no phycobilin signal compared to strain MBIC11017 (Fig. 1b). Furthermore, spectra for strain CCMEE5410 did not demonstrate a significant peak at 680 nm where one would expect the Chl *a* Q_y absorption peak, although this is not surprising given the low level of Chl *a* which has been reported for *A. marina* (Miller *et al.*, 2005). In light of this, we assign wavelengths between 590 and 680 nm from whole cell spectra to be due to phycobilin absorption (Fig. 1b) compared to Schiller *et al.* (1997) where previously the phycobilin band was assigned between 580 and 660 nm. The Chl *d* Q_y absorption peak lay between 715 and 717 nm after light scattering correction for both strains.

To determine how light intensity and temperature affected pigment composition, cultures were grown under varying light intensities and at two different temperatures. Strains MBIC11017 and CCMEE5410 both demonstrated a pigmentation adaptation mechanism. Cultures began to adapt after 2 days and the adaptation was prominent after 4 days. To compare spectra, the absorbance data were normalized at 580 nm, which revealed the relative changes in the most prominent pigmentation peaks in whole cell spectra: those of Chl *d* and the PBPs. In MBIC11017 cultures grown at 28 °C, relative absorbance spectra showed that an increase in light intensity lowered the absorbance of both Chl *d* and the phycobilins of intact cells (Fig. 1a and b). Similarly, relative absorbance spectra in strain CCMEE5410 showed a gradual decrease in the Chl *d* peak at 716 nm with respect to

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

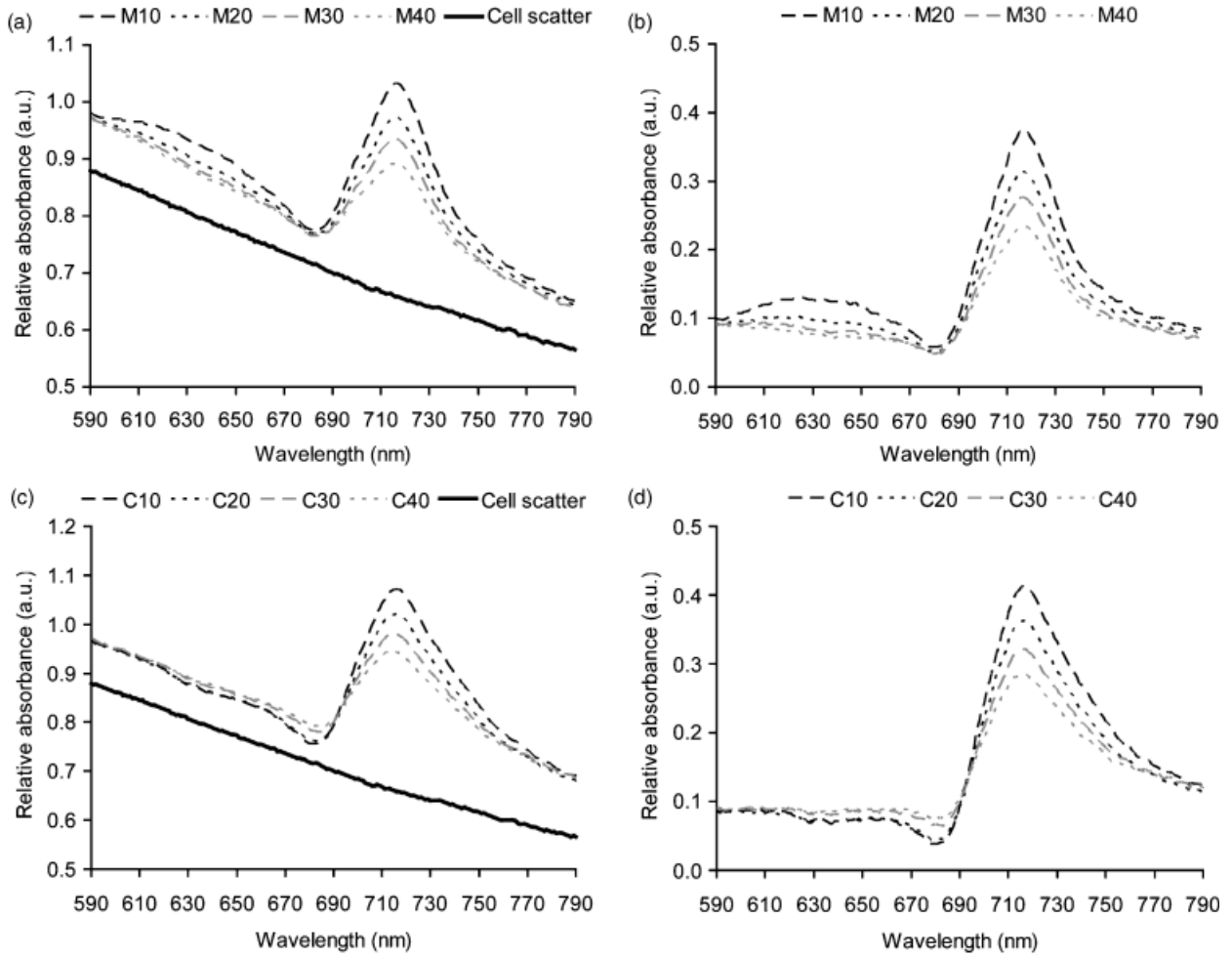


Fig. 1. Absorbance spectra of strains MBIC11017 (M) and CCMEE5410 (C) grown at 28 °C under continuous white light at 10, 20, 30 and 40 μE m^{-2} s^{-1}, 10 days after inoculation after light adaptation. Whole cell spectra were normalized at 580 nm (a and c) and the corresponding light scattering corrected spectra were obtained after the subtraction of cell scatter data (b and d). The figure legends indicate strain and light intensity.

increasing light intensity (Fig. 1c and d). The same results were obtained at 23 °C and under light/dark conditions at 28 °C but cells were slower to adapt. Once light adaptation had occurred, the characteristic gradation of spectra remained, irrespective of growth phase for both *Acaryochloris* strains (Fig. 1a and b).

Strains MBIC11017 and CCMEE5410 were nonaxenic due to the presence of a persistent heterotrophic bacterium, which grew on Luria–Bertani plates. Data were collected after repeated streak-plating and culturing. To check that this would not have an effect on the results reported in this study, growth properties of *A. marina* were monitored using both OD and flow cytometry. Nonshaken cultures were observed to have a fluctuating growth rate as opposed to shaken cultures, which showed a steadily increasing growth rate with age. The same pigment adaptation was observed under both growth regimes. From this we concluded that the pigment adaptation was independent of *Acaryochloris* growth rate.

Fluorescence measurements

Emission and excitation spectra for both strains were recorded at the end of each absorbance time course to investigate whether the fluorescence dynamics of the whole cells were affected by growth at different light intensities. Previously, Petrášek *et al.* (2005) showed that the intensity of PBP fluorescence is highly dependent on the excitation wavelength. This was confirmed by preliminary fluorescence optimization experiments. In light of this, emission spectra were measured by exciting at 630 nm to demonstrate relative fluorescence intensities of PBPs and Chl *d* in the two strains.

Strain MBIC11017 showed characteristic peaks around 640–650 and 717 nm, which are attributed to PBPs and

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

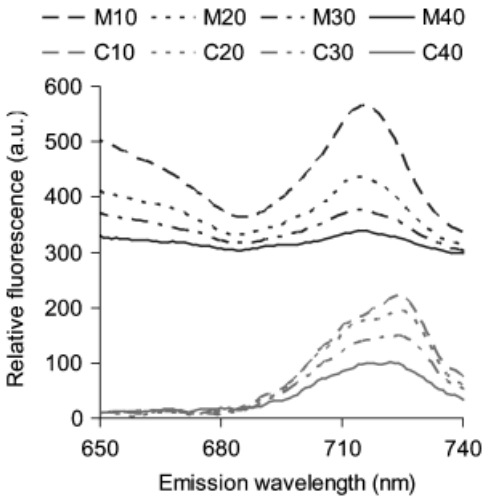


Fig. 2. Emission spectra obtained by excitation at 630 nm for strains MBIC11017 (M) and CCMEE5410 (C) grown at 28 °C under continuous white light at 10, 20, 30 and 40 $\mu E\, m^{-2}\, s^{-1}$. Spectra were recorded at room temperature 10 days after inoculation after light adaptation and are normalized relative to the absorbance at 800 nm of the same samples. The figure legend indicates strain and light intensity.

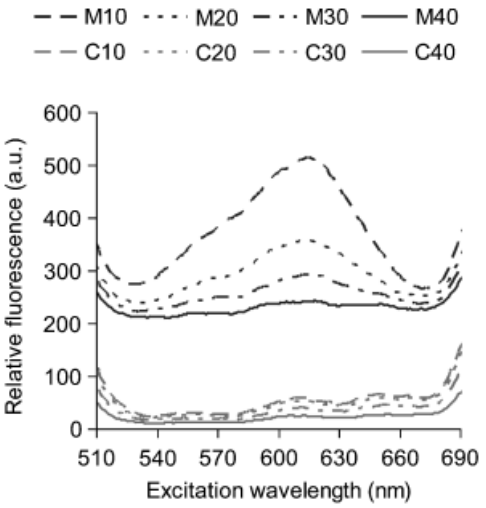


Fig. 3. Excitation spectra obtained for strains MBIC11017 (M) and CCMEE5410 (C) with emission wavelengths at 717 and 725 nm, respectively. Cultures were grown at 28 °C under continuous white light at 10, 20, 30 and 40 $\mu E\, m^{-2}\, s^{-1}$ and recorded at room temperature 10 days after inoculation after light adaptation. Spectra have been normalized relative to the absorbance at 800 nm of the same samples. The figure legend indicates strain and light intensity.

Chl *d*, respectively. A gradual decrease in these peak wavelength bands was observed with respect to increasing growth light intensity, indicating the loss of PBPs and Chl *d* at higher light intensities (Fig. 2). Strain CCMEE5410 showed a similar decrease in the Chl *d* peak at 725 nm with respect to an increase in growth light intensity, and again no peak corresponding to PBPs was observed.

Excitation spectra were recorded by collecting the emission at 717 and 725 nm for strains MBIC11017 and CCMEE5410, respectively (Fig. 3), i.e. the emission wavelengths assigned to Chl *d* for each strain. The excitation spectra were measured between 510 and 690 nm to show the degree of excitation energy transfer from PBPs to Chl *d* in strain MBIC11017, which demonstrated a decrease in signal intensity at higher growth irradiances. For strain CCMEE5410 the excitation spectra of all samples resembled the spectrum for strain MBIC11017 at 40 $\mu E\, m^{-2}\, s^{-1}$, though there was a gradually decreased signal over the growth irradiance range used. The emission spectra of strain CCMEE5410 indicates that either Chl *d* or some other pigment is able to absorb light at 630 nm and to contribute to Chl *d* fluorescence emission in the absence of PBPs (Fig. 2).

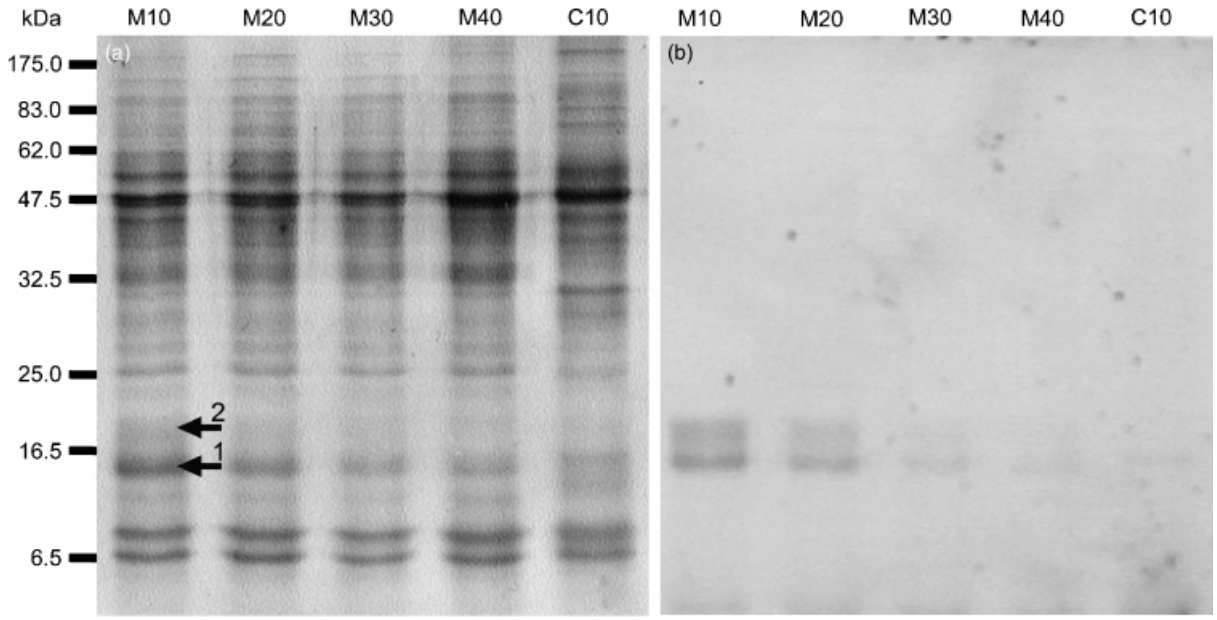
Protein analysis

SDS-PAGE of soluble protein from whole cell samples (grown at 28 °C under continuous white light at 10, 20, 30 and 40 $\mu E\, m^{-2}\, s^{-1}$), revealed a decrease in the intensity of the chromophore bands, visible under UV, corresponding to PC and APC (Marquardt *et al.*, 1997) (Fig. 4b) for strain MBIC11017 with respect to increasing growth irradiance. The crude soluble protein from strain CCMEE5410 cells grown under the same conditions appeared to completely lack chromophore bands corresponding to PC and APC under UV illumination (Fig. 4a and b, lane C10). Due to the high degree of similarity between the lanes (corresponding to different growth light intensities) for strain CCMEE5410, only the lane for cells grown at 10 $\mu E\, m^{-2}\, s^{-1}$ is shown (lane C10).

LC-MS/MS analysis of each extracted protein band and subsequent MASCOT searches identified the proteins present in the chromophore bands for strain MBIC11017 as being the α and β subunits of phycocyanin (Fig. 4c). Phycocyanin peptides identified here match those found in several cyanobacteria. Although the *A. marina* genome has not yet been annotated, sequence data for MBIC11017 are available (http://genomes.tgen.org/) and all peptides identified here are predicted in the *A. marina* genome.

Bands that were removed from the corresponding 10 to 20 kDa size region in strain CCMEE5410 and analyzed using LC-MS/MS did not reveal any evidence of peptides corresponding to PC or APC proteins. Indeed, the peptides which were recovered matched a DNA-binding stress protein (*Synechococcus* sp., strain PCC6301, Q5N285_SYNP6), thioredoxin (*Synechococcus* sp., strain WH8102, Q7U898_SYNPX) and photosystem I reaction centre subunit II (photosystem I 16 kDa polypeptide; PSI-D, *Synechococcus elongatus*, PSAD_SYNEL).

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved



(c) Peptides identified by LC-MS/MS

(1) C-phycocyanin 2 α chain (*Calothrix* sp., accession number: B28539)

```
  1 MKTPLTEAVA TADSQGRFLS STELQVAFGR FRQASASLDA AKALSSKANS
 51 LAQGAVNAVY QKFPYTTQMQ GKNFASDQRG KDKCARDIGY YIRIVTYCLV
101 AGGTGPLDDY LIGGLAEINR TFDLSPSWYV EALKYIKANH GLSGDPAVEA
151 NSYIDYAINA LS
```

(2) Phycocyanin β subunit (*A. marina*, accession number: Q5GJ62)

```
  1 KLLSVVSDGN KRMDVVNRIT SNASTIVANA ARDLFEEQPS LIQPGGNAYT
 51 HRRMAACLRD MEIILRYVTY AIFAGDASIL EDRCLNGLKQ TYQTLGVPTK
101 SMSLSVSKMR DAALEIASDP NGVTQGDCSS LISEVSDYFD LAARAVG
```

Fig. 4. (a) Coomassie-stained SDS-PAGE gel and (b) UV image of the same gel of soluble protein extracts from MBIC11017 (M) and CCMEE5410 (C) cultures. Cells were grown at 28 °C under continuous white light at 10, 20, 30 and 40 μE m^{-2} s^{-1}. Lanes are labeled by strain and light intensity. Peptides identified from LC-MS/MS are highlighted in bold for the bands indicated in the SDS-PAGE gel with arrowheads.

This result was further strengthened by Western blotting (Fig. 5), which shows that in MBIC11017 the amount of APC and PC is reduced under higher light conditions and they are absent in CCMEE5410. The Western blot showed clear bands of ~58 kDa in all samples (it is very faint in lane M10, but present). These bands are thought to be due to the nonspecific binding of the antibody to an unknown protein, as no corresponding chromophore band was visible under UV.

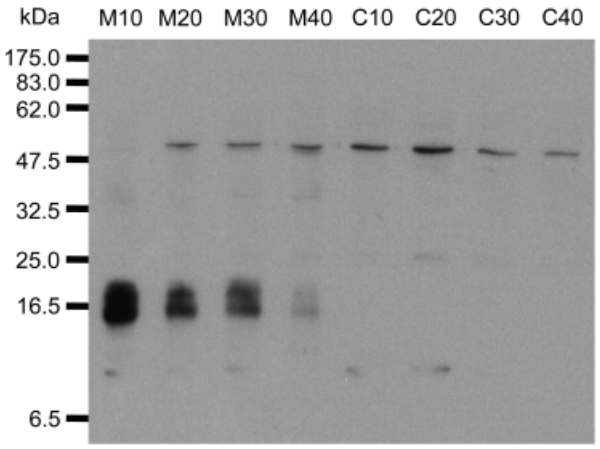


Fig. 5. Western blot of soluble protein extracts from MBIC11017 (M) and CCMEE5410 (C) cultures. Cells were grown at 28 °C under continuous white light at 10, 20, 30 and 40 μE m^{-2} s^{-1}. Lanes are labeled by strain and light intensity.

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

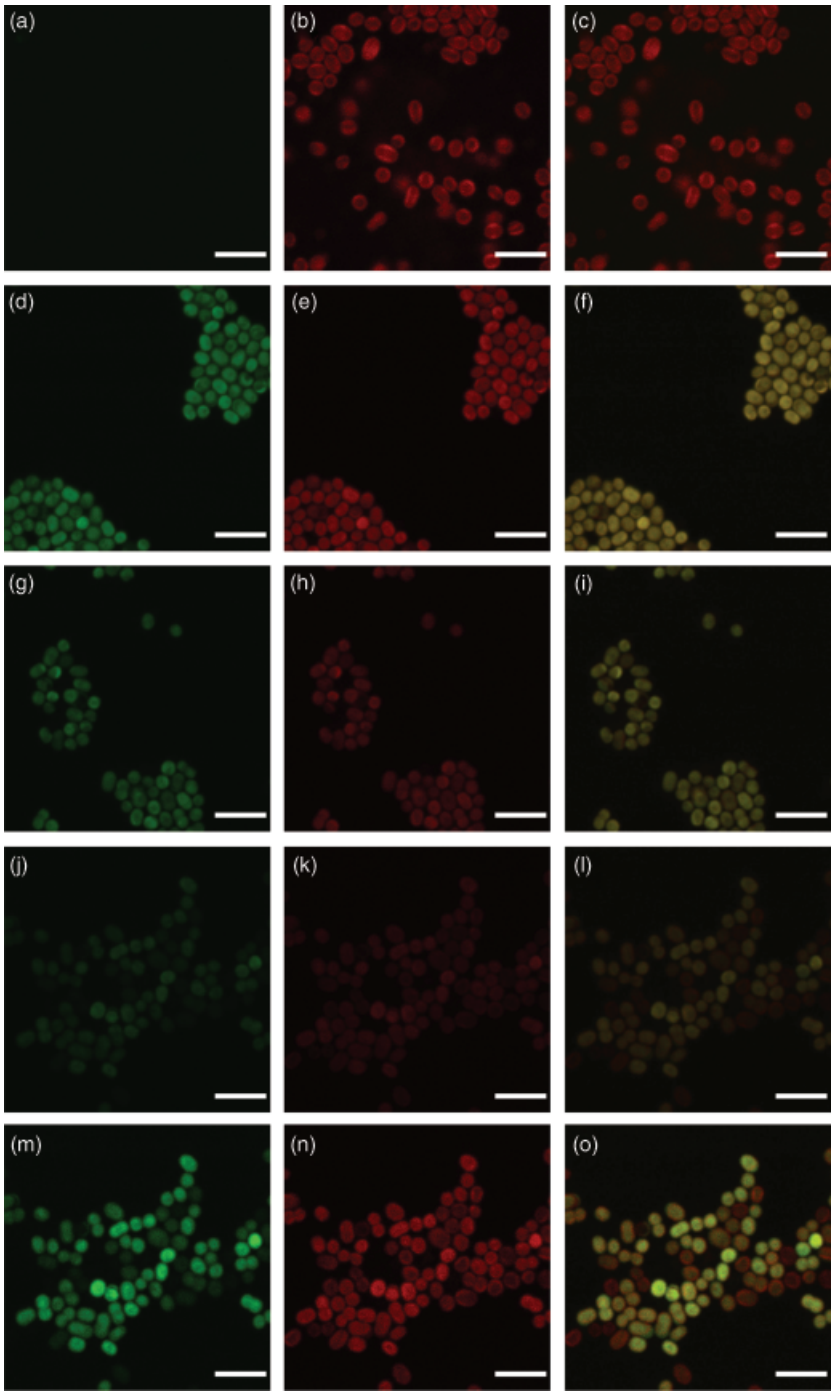


Fig. 6. LSCM images of *A. marina* strains MBIC11017 and CCMEE5410. (a–c) CCMEE5410 cells. MBIC11017 cells grown at 25 μE m^{-2} s^{-1} (d–f), 50 μE m^{-2} s^{-1} (g–i) and 100 μE m^{-2} s^{-1} (j–l). Images (m–o) show the same cells as in (j–l) but with increased gains in the PBP and Chl *d* channels. MBIC11017 cells were grown at 28 °C, while being shaken and harvested for imaging 136 h after inoculation. Images in green (a, d, g, j, m) and red (b, e, h, k, n) with composite images (c, f, i, l, o), respectively, from combining the PBP and Chl *d* channels. Scale bar, 6.3 μm.

Laser scanning focal microscopy (LSCM)

LSCM was an extremely useful method to examine the pigment composition of *A. marina* grown under different light regimes. Confocal images showed that CCMEE5410 cells had no signal in the PBP channel (Fig. 6a). In comparison, for MBIC11017 cells grown at 100 μE m^{-2} s^{-1}, even a low PBP content is detectable (Fig. 6j), suggesting the complete absence of PBPs in strain CCMEE5410. Comparison of Fig. 6a and b shows that fluorescence from Chl *d* does not contribute to the PBP channel signal.

At the lower light intensity of 25 μE m^{-2} s^{-1}, shaken MBIC11017 cells showed strong signals in the PBP (Fig. 6d) and Chl *d* (Fig. 6e) channels. The decrease in fluorescence and abundance of PBPs and Chl *d* corresponds to the absorbance and fluorescence data recorded for these

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

cultures as observed for the nonshaken cultures (Figs 1–3). Under increased gains, shaken MBIC11017 cells grown at 100 μE m^{-2} s^{-1}, contained a third subpopulation (Fig. 6o). This heterogeneous pigment composition is most likely attributed to natural shading, as it was observed in both shaken and nonshaken cells grown under higher light. The phenomenon of excitation energy transfer from PBPs to Chl *d* was particularly evident in these cultures.

We conclude that MBIC11017 cells exposed to higher light intensities adapt by losing their PBPs as they no longer require them to increase the efficiency of light-harvesting (Boichenko *et al.*, 2000; Petrášek *et al.*, 2005). This is consistent with the predominantly low light, yet fluctuating, environment in which *A. marina* is found. In contrast, strain CCMEE5410 lacks PBPs, which implies that PBPs are not required for survival when *Acaryochloris* are growing endolithically, in the shallow water and high light conditions of the Salton Sea.

The Salton Sea is a highly eutrophic artificial lake maintained by agricultural runoff (http://www.saltonsea.ca.gov/). The *Acaryochloris* strain which was isolated from here was growing as an epilithic mat and cultured under high light conditions of 50 μE m^{-2} s^{-1}, presumably reflecting the environmental conditions in which it was discovered (Miller *et al.*, 2005). Our results indicate that strain MBIC11017 loses its PBPs at light intensities greater than or equal to 40 μE m^{-2} s^{-1}. It is likely that strain CCMEE5410 is a derivative of a PBP-producing strain akin to strain MBIC11017, which may have been inadvertently introduced into the Salton Sea from marine fish stocks (Ritchie, 2006). Prolonged exposure to high light intensities may have caused the ancestral strain to lose the ability to produce PBPs. Sequence data should indicate patterns of ancestry and whether the genes encoding PBPs are still present but nonfunctional in strain CCMEE5410 or whether they are missing altogether.

Acknowledgements

This research was funded by the EPSRC. We would like to acknowledge Michelle Wood for the kind gift of strain CCMEE5410 and Wolfgang Hess for strain MBIC11017. The MS analysis was a contribution from the Edinburgh Protein Interaction Centre funded by the Wellcome Trust. ALW would like to thank the Royal Society for a Royal Society BP Dorothy Hodgkin Research Fellowship. We would like to thank Antony Holmes for several useful data processing programs and Professor Alison Rodger for useful discussions on fluorescence spectroscopy.

References

Aitken A & Learmonth M (2002) Protein identification by in-gel digestion and mass spectrometric analysis. *Mol Biotechnol* **20**: 95–97.

Akimoto S, Murakami A, Yokono M, Koyama K, Tsuchiya T, Miyashita H, Yamazaki I & Mimuro M (2006) Fluorescence properties of the chlorophyll *d*-dominated cyanobacterium *Acarychloris* sp. strain Awaji. *J Photochem Photobiol A: Chem* **178**: 122–129.

Akiyama M, Miyashita H, Kise H, Watanabe T, Miyachi S & Kobayashi M (2001) Detection of chlorophyll *d*′ and pheophytin a in a chlorophyll *d*-dominating oxygenic photosynthetic prokaryote *Acaryochloris marina*. *Anal Sci* **17**: 205–208.

Berkelman TR & Lagarias JC (1986) Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal Biochem* **156**: 194–201.

Boichenko VA, Klimov VV, Miyashita H & Miyachi S (2000) Functional characteristics of chlorophyll *d*-predominating photosynthetic apparatus in intact cells of *Acaryochloris marina*. *Photosynth Res* **65**: 269–277.

Chen M, Bibby TS, Nield J, Larkum A & Barber J (2005a) Iron deficiency induces a chlorophyll *d*-binding Pcb antenna system around Photosystem I in *Acaryochloris marina*. *Biochim Biophys Acta* **1708**: 367–374.

Chen M, Telfer A, Lin S, Pascal A, Larkum AW, Barber J & Blankenship RE (2005b) The nature of the photosystem II reaction centre in the chlorophyll *d*-containing prokaryote, *Acaryochloris marina*. *Photochem Photobiol Sci* **4**: 1060–1064.

de los Ríos A, Grube M, Sancho LG & Ascaso C (2007) Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS Microbiol Ecol* **59**: 386–395.

Eng JK, McCormack AL & Yates JR III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**: 976–989.

Engelmann TW (1883) Farbe und assimilation. *Bot Z* **41**: 1–29.

Hu Q, Miyashita H, Iwasaki II, Kurano N, Miyachi S, Iwaki M & Itoh S (1998) A photosystem I reaction center driven by chlorophyll *d* in oxygenic photosynthesis. *Proc Natl Acad Sci USA* **95**: 13319–13323.

Hu Q, Marquardt J, Iwasaki I, Miyashita H, Kurano N, Mörschel E & Miyachi S (1999) Molecular structure, localization and function of biliproteins in the chlorophyll *a*/*d* containing oxygenic photosynthetic prokaryote *Acaryochloris marina*. *Biochim Biophys Acta* **1412**: 250–261.

Kobayashi M, Watanabe S, Gotoh T *et al.* (2005) Minor but key chlorophylls in photosystem II. *Photosynth Res* **84**: 201–207.

Kühl M, Chen M, Ralph PJ, Schreiber U & Larkum AW (2005) Ecology: a niche for cyanobacteria containing chlorophyll *d*. *Nature* **433**: 820.

Kumazaki S, Abiko K, Ikegami I, Iwaki M & Itoh S (2002) Energy equilibration and primary charge separation in chlorophyll *d*-based photosystem I reaction center isolated from *Acaryochloris marina*. *FEBS Lett* **530**: 153–157.

Larkum AW & Kühl M (2005) Chlorophyll *d*: the puzzle resolved. *Trends Plant Sci* **10**: 355–357.

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Marquardt J, Senger H, Miyashita H, Miyachi S & Mörschel E (1997) Isolation and characterization of biliprotein aggregates from *Acaryochloris marina*, a *Prochloron*-like prokaryote containing mainly chlorophyll *d*. *FEBS Lett* **410**: 428–432.

Miller SR, Augustine S, Olson TL, Blankenship RE, Selker J & Wood AM (2005) Discovery of a free-living chlorophyll *d*-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc Natl Acad Sci USA* **102**: 850–855.

Mimuro M, Hirayama K, Uezono K, Miyashita H & Miyachi S (2000) Uphill energy transfer in a chlorophyll *d*-dominating oxygenic photosynthetic prokaryote, *Acaryochloris marina*. *Biochim Biophys Acta* **1456**: 27–34.

Mimuro M, Akimoto S, Gotoh T, Yokono M, Akiyama M, Tsuchiya T, Miyashita H, Kobayashi M & Yamazaki I (2004) Identification of the primary electron donor in PS II of the Chl *d*-dominated cyanobacterium *Acaryochloris marina*. *FEBS Lett* **556**: 95–98.

Miyashita H, Ikemoto H, Kurano N, Adachi K, Chihara M & Miyachi S (1996) Chlorophyll *d* as a major pigment. *Nature* **383**: 402.

Miyashita H, Adachi K, Norihide K, Ikemoto H, Chihara M & Miyachi S (1997) Pigment composition of a novel oxygenic photosynthetic prokaryote containing chlorophyll *d* as the major chlorophyll. *Plant Cell Physiol* **38**: 274–281.

Murakami A, Miyashita H, Iseki M, Adachi K & Mimuro M (2004) Chlorophyll *d* in an epiphytic cyanobacterium of red algae. *Science* **303**: 1633.

Ohkubo S, Miyashita H, Murakami A, Takeyama H, Tsuchiya T & Mimuro M (2006) Molecular detection of epiphytic *Acaryochloris* spp. on marine macroalgae. *Appl Environ Microbiol* **72**: 7912–7915.

Oltmanns F (1892) Über die Kultur und Lebensbedingungen der Meersalgen. *Jahrb Wiss Botanik* **23**: 349–440.

Petrášek Z, Schmitt FJ, Theiss C, Huyer J, Chen M, Larkum A, Eichler HJ, Kemnitz K & Eckert HJ (2005) Excitation energy transfer from phycobiliprotein to chlorophyll *d* in intact cells of *Acaryochloris marina* studied by time- and wavelength-resolved fluorescence spectroscopy. *Photochem Photobiol Sci* **4**: 1016–1022.

Ritchie RJ (2006) Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res* **89**: 27–41.

Savitzky A & Golay MJE (1964) Smoothing and differentiation of data by simplified least square procedures. *Anal Chem* **36**: 1627–1639.

Schiller H, Senger H, Miyashita H, Miyachi S & Dau H (1997) Light-harvesting in *Acaryochloris marina* – spectroscopic characterization of a chlorophyll *d*-dominated photosynthetic antenna system. *FEBS Lett* **410**: 433–436.

Swingley WD, Hohmann-Marriott MF, Le Olson T & Blankenship RE (2005) Effect of iron on growth and ultrastructure of *Acaryochloris marina*. *Appl Environ Microbiol* **71**: 8606–8610.

Wyman M, Gregory RPF & Carr NG (1985) Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**: 818–820.

© 2007 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Bibliography

S. T. Abedon. *Bacteriophage Ecology. Population Growth, Evolution and Impact of Bacterial Viruses*. Cambridge University Press, 2008.

S. T. Abedon, T. D. Herschler, and D. Stopar. Bacteriophage latent-period evolution as a response to resource availability. *Appl Environ Microbiol*, 67(9):4233–4241, 2001.

S. T. Abedon, P. Hyman, and C. Thomas. Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. *Appl Environ Microbiol*, 69(12):7499–7506, 2003.

A. Aitken and M. Learmonth. Protein identification by in-gel digestion and mass spectrometric analysis. *Mol Biotechnol*, 20(1):95–97, 2002.

B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell, Fourth Edition*. Garland, 2002.

L. M. Allen, M. R. Hodskinson, and Sayers J. R. Active site substitutions delineate distinct classes of eubacterial flap endonuclease. *Biochem J*, 418(2):285–292, 2009.

L. A. Allison, L. D. Simon, and P. Maliga. Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J*, 15 (11):2802–2809, 1996.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, 1990.

S. Amemiya, J. Guo, H. Xiong, and D. A. Gross. Biological applications of scanning electrochemical microscopy: chemical imaging of single living cells and beyond. *Anal Bioanal Chem*, 386(3):458–471, 2006.

A. F. Andersson and J. F. Banfield. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, 320:1047–1050, 2008.

A. C. Baker, V. J. Goddard, J. Davy, D. C. Schroeder, D. G. Adams, and W. H. Wilson. Identification of a diagnostic marker to detect freshwater cyanophages of filamentous cyanobacteria. *Appl Environ Microbiol*, 72: 5713–5719, 2006.

R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315:1709–1712, 2007.

Ø. Bergh, K. Y. Børsheim, G. Bratbak, and M. Heldal. High abundance of viruses found in aquatic environments. *Nature*, 340(6233):467–468, 1989.

T. R. Berkelman and J. C. Lagarias. Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal Biochem*, 156(1):194–201, 1986.

J. Besemer and M. Borodovsky. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res*, 27(19):3911–3920, 1999.

Applied Biosystems. *Creating standard curves with genomic DNA or plasmid DNA templates for use in quantitative PCR*, 4371090 revision a edition, 2003.

V. A. Boichenko, V. V. Klimov, H. Miyashita, and S. Miyachi. Functional characteristics of chlorophyll *d*-predominating photosynthetic apparatus in intact cells of *Acaryochloris marina*. *Photosynth Res*, 65(3):269–277, 2000.

A. Bolden, G. Pedrali-Noy, and A. Weissbach. DNA polymerase of mitochondria is a gamma-polymerase. *J Biol Chem*, 252(10):3351–3356, 1977.

J. G. Bragg and S. W. Chisholm. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS ONE*, 3(10):e3550, 2008.

D. K. Braithwaite and J. Ito. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res*, 21(4):787–802, 1993.

S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891):960–964, 2008.

C. P. D. Brussaard. Viral control of phytoplankton populations - a review. *J Eukaryot Microbiol*, 51(2):125–138, 2004.

H. Brüssow and F. Desiere. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol Microbiol*, 39(2):213–222, 2001.

M. Buchwald, H. Murialdo, and L. Siminovitch. The morphogenesis of bacteriophage lambda. II. Identification of the principal structural proteins. *Virology*, 42(2):390–400, 1970.

S. A. Bustin. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol*, 25: 169–193, 2000.

D. Campbell, V. Hurry, A. K. Clarke, P. Gustafsson, and G. Öquist. Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation. *Microbiol Mol Biol Rev*, 62(3):667–683, 1998.

C. Canchaya, G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brüssow. Phage as agents of lateral gene transfer. *Curr Opin Microbiol*, 6(4):417–424, 2003.

T. J. Carver, K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill. ACT: the Artemis Comparison Tool. *Bioinformatics*, 21(16):3422–3423, 2005.

C. E. Catalano. The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. *Cell Mol Life Sci*, 57(1):128–148, 2000.

C. E. Catalano, editor. *Viral genome packaging machines: genetics, structure, and mechanism*. Kluwer Academic/Plenum Publishers, 2005.

Y. W. Chan, A. Nenninger, S. J. Clokie, N. H. Mann, D. J. Scanlan, A. L. Whitworth, and M. R. J. Clokie. Pigment composition and adaptation in free-living and symbiotic strains of *Acaryochloris marina*. *FEMS Microbiol Ecol*, 61(1):65–73, 2007.

F. Chen and J. Lu. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol*, 68(5):2589–2594, 2002.

M. Chen, R. G. Quinnell, and A. W. Larkum. The major light-harvesting pigment protein of *Acaryochloris marina*. *FEBS Lett*, 514(2-3):149–152, 2002.

M. Chen, A. Telfer, S. Lin, A. Pascal, A. W. Larkum, J. Barber, and R. E. Blankenship. The nature of the photosystem II reaction centre in the chlorophyll *d*-containing prokaryote, *Acaryochloris marina*. *Photochem Photobiol Sci*, 4(12):1060–1064, 2005.

M. Chen, M. Floetenmeyer, and T. S. Bibby. Supramolecular organization of phycobiliproteins in the chlorophyll *d*-containing cyanobacterium *Acaryochloris marina*. *FEBS Lett*, Epub ahead of print:pp. unknown, 2009.

B. S. Chevalier and B. L. Stoddard. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res*, 29(18):3757–3774, 2001.

S. W. Chisholm, S. L. Frankel, R. Goericke, R. J. Olson, B. Palenik, J. B. Waterbury, L. West-Johnsrud, and E. R. Zettler. *Prochlorococcus marinus* nov. gen. nov. sp. an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch Microbiol*, 157:297–300, 1992.

A. Chopin, A. Bolotin, A. Sorokin, S. D. Ehrlich, and M. Chopin. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res*, 29(3): 644–651, 2001.

H. Christensen, P. Kuhnert, J. E. Olsen, and M. Bisgaard. Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *Int J Syst Evol Microbiol*, 54: 1601–1609, 2004.

M. Ciobanu, D. E. Taylor Jr., J. P. Wilburn, and D. E. Cliffel. Glucose and lactate biosensors for scanning electrochemical microscopy imaging of single live cells. *Anal Chem*, 80:2717–2727, 2008.

M. R. Clokie and A. M. Kropinski, editors. *Bacteriophages: Methods and Protocols (Methods in Molecular Biology)*. Humana Press, 2009.

M. R. Clokie and N. H. Mann. Marine cyanophages and light. *Environ Microbiol*, 8(12):2074–2082, 2006.

M. R. Clokie, J. Shan, S. Bailey, Y. Jia, H. M. Krisch, S. West, and N. H.

Mann. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol*, 8(5):827–835, 2006a.

M. R. Clokie, K. Thalassinos, P. Boulanger, S. E. Slade, S. Stoilova-McPhie, M. Cane, J. H. Scrivens, and N. H. Mann. A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. *Microbiology*, 154(Pt 6):1775–1782, 2008.

M. R. J. Clokie. Quantification of host and phage mRNA expression during infection using real-time PCR. *Methods Mol Biol*, 502:177–191, 2009.

M. R. J. Clokie, A. D. Millard, J. Y. Mehta, and N. H. Mann. Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. *J Mar Biol Ass UK*, 86 (3):499–505, 2006b.

J. G. Cobley, A. C. Clark, S. Weerasurya, F. A. Queseda, J. Y. Xiao, N. Bandrapali, I. D'Silva, M. Thounaojam, J. F. Oda, T. Sumiyoshi, and M. H. Chu. CpeR is an activator required for expression of the phycoerythrin operon (*cpeBA*) in the cyanobacterium *Fremyella diplosiphon* and is encoded in the phycoerythrin linker-polypeptide operon (*cpeCDESTR*). *Mol Microbiol*, 44(6):1517–1531, 2002.

C. Condon and H. Putzer. The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res*, 30(24):5339–5346, 2002.

J. G. Day. Cryopreservation of microalgae and cyanobacteria. *Methods Mol Biol*, 368:141–151, 2007.

A. de los Ríos, M. Grube, L. G. Sancho, and C. Ascaso. Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS Microbiol Ecol*, 59(2):386–395, 2007.

E. F. DeLong, C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm,

and D. M. Karl. Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, 311(5760):496–503, 2006.

M. P. Deutscher. Synthesis and functions of the -C-C-A terminus of transfer RNA. *Prog Nucleic Acid Res Mol Biol*, 13:51–92, 1973.

M. P. Deutscher and C. W. Marlor. Purification and characterization of *Escherichia coli* RNase T. *J Biol Chem*, 260(11):7067–7071, 1985.

M. P. Deutscher, C. W. Marlor, and R. Zaniewski. Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA. *Proc Natl Acad Sci U S A*, 81(14):4290–4293, 1984.

M. P. Deutscher, C. W. Marlor, and R. Zaniewski. RNase T is responsible for the end-turnover of tRNA in *Escherichia coli*. *Proc Natl Acad Sci U S A*, 82(19):6427–6430, 1985.

U. Dorigo, S. Jacquet, and Humbert J. F. Cyanophage diversity, inferred from g20 gene analyses, in the largest natural lake in France, Lake Bourget. *Appl Environ Microbiol*, 70:1017–1022, 2004.

Z. Duxbury, M. Schliep, R. J. Ritchie, A. W. D. Larkum, and M. Chen. Chromatic photoacclimation extends utilisable photosynthetically active radiation in the chlorophyll *d*-containing cyanobacterium, *Acaryochloris marina*. *Photosynth Res*, Epub ahead of print:pp. unknown, 2009.

H. Elmsford, J. Riguady, and S. P. Klesney, editors. *Nomenclature of Organic Chemistry, Sections A, B, C, D, E, F, and H (Copyright 1979 IUPAC)*. Pergamon Press, 1979.

J.K. Eng, A.L. McCormack, and J.R. Yates III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*, 5:976–989, 1994.

T.W. Engelmann. Farbe und Assimilation. *Bot Z*, 41:1–29, 1883.

S. Eriksson, B. Munch-Petersen, K. Johansson, and H. Eklund. Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol Life Sci*, 59(8):1327–1346, 2002.

M. D. Ermolaeva, H. G. Khalak, O. White, H. O. Smith, and S. L. Salzberg. Prediction of transcription terminators in bacterial genomes. *J Mol Biol*, 301:27–33, 2000.

M. Falkenberg, N. G. Larsson, and C. M. Gustafsson. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem*, 76:679–699, 2007.

J. Filee and P. Forterre. Viral proteins functioning in organelles: a cryptic origin? *Trends Microbiol*, 13(11):510–513, 2005.

J. Filee, P. Forterre, T. Sen-Lin, and J. Laurent. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J Mol Evol*, 54(6):763–773, 2002.

J. Filée, F. Tétart, C. A. Suttle, and Krisch H. M. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc Natl Acad Sci U S A*, 102:12471–12476, 2005.

L. C. Fortier, A. Bransi, and S. Moineau. Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of *Lactococcus lactis*. *J Bacteriol*, 188(17):6101–6114, 2006.

J. S. Fraser, Z. Yu, K. L. Maxwell, and A. R. Davidson. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol*, 359: 496–507, 2006.

N. J. Fuller, W. H. Wilson, I. R. Joint, and N. H. Mann. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. *Appl Environ Microbiol*, 64(6):2051–2060, 1998.

N. J. Fuller, D. Marie, F. Partensky, D. Vaulot, A. F. Post, and D. J. Scanlan. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol*, 69(5):2430–2443, 2003.

J. L. Gardy, C. Spencer, K. Wang, M. Ester, G. E. Tusnady, I. Simon, S. Hua, K. deFays, C. Lambert, K. Nakai, and F. S. Brinkman. PSORT-B: improving protein subcellular localization prediction for gram-negative bacteria. *Nucleic Acids Res*, 31(13):3613–3617, 2003.

D. M. Gates. *Biophysical Ecology*. Springer-Verlag, New York, 1980.

F. S. Gimble. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett*, 185(2):99–107, 2000.

A. N. Glazer. Phycobilisomes: structure and dynamics. *Annu Rev Microbiol*, 36:173–198, 1982.

R. Goericke and N. A. Welschmeyer. The marine prochlorophyte *Prochlorococcus* contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea. *Deep-Sea Res Part I-Oceanographic Res Pap*, 40:2283–2294, 1993.

F. Goh, M. A. Allen, S. Leuko, T. Kawaguchi, A. W. Decho, B. P. Burns, and B. A. Neilan. Determining the specific microbial populations and their spatial distribution within the stromatolite ecosystem of Shark Bay. *Isme J*, 3(4):383–396, 2009.

J. M. Gonzalez and M. A. Moran. Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. *Appl Environ Microbiol*, 63(11):4237–4242, 1997.

M. W. Gray, G. Burger, and B. F. Lang. Mitochondrial evolution. *Science*, 283(5407):1476–1481, 1999.

I. Grissa, G. Vergnaud, and C. Pourcel. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, 35(Web Server issue):W52–W57, 2007a.

I. Grissa, G. Vergnaud, and C. Pourcel. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 8:172, 2007b.

N. Hall. Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol*, 209:1518–1525, 2007.

E. H. Hankin. L'action bactericide des eaux de la Jumna et du Gange sur le vibrion du cholera. *Ann Inst Pasteur*, 10:511, 1896.

F. Harada, T. Nakano, T. Kohno, S. Mohan, H. Taniguchi, and K. Sano. RNA-dependent DNA polymerase (rt) activity of bacterial DNA polymerases. *Bull Osaka Med Coll*, 51(1):35–41, 2005.

J. J. Harrington and M. R. Lieber. The characterisation of a mammalian DNA structure-specific endonuclease. *EMBO J*, 13(5):1235–1246, 1994.

P. V. Harris, O. M. Mazina, E. A. Leonhardt, R. B. Case, J. B. Boyd, and K. C. Burtis. Molecular cloning of *Drosophila* mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. *Mol Cell Biol*, 16(10):5764–5771, 1996.

Heinz Walz GmbH. *Phytoplankton analyzer. PHYTO-PAM and Phyto-Win Software V 1.45. System components and principles of operation*. Eichenring 6, 91090 Effeltrich, Germany, 2 edition, July 2003.

F. L. Hellweger. Carrying photosynthesis genes increases ecological fitness of cyanophage *in silico*. *Environ Microbiol*, 11:1386–1394, 2009.

R. W. Hendrix, M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. Evolutionary relationships among diverse bacteriophages and

prophages: All the worlds a phage. *Proc Natl Acad Sci U S A*, 96:2192–2197, 1999.

W. R. Hess and T. Borner. Organellar RNA polymerases of higher plants. *Int Rev Cytol*, 190:1–59, 1999.

R. Higuchi, C. Fockler, G. Dollinger, and R. Watson. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nature Biotechnology*, 11(9):1026–1030, 1993.

H. C. Hollingsworth and N. G. Nossal. Bacteriophage T4 encodes an RNase H which removes RNA primers made by the T4 DNA replication system *in vitro*. *J Biol Chem*, 266(3):1888–1897, 1991.

A. S. Holt and H. V. Morley. Chlorophyll *a* with the 9-vinyl group replaced by a formyl group. *Can J Chem*, 37:507–514, 1959.

A.S. Holt. Further evidence of the relation between 2-desvinyl-2-formylchlorophyll *a* and chlorophyll *d*. *Can J Bot*, 39:327–331, 1961.

Q. Hu, H. Miyashita, I. I. Iwasaki, N. Kurano, S. Miyachi, M. Iwaki, and S. Itoh. A photosystem I reaction center driven by chlorophyll *d* in oxygenic photosynthesis. *Proc Natl Acad Sci U S A*, 95(22):13319–13323, 1998.

Q. Hu, J. Marquardt, I. Iwasaki, H. Miyashita, N. Kurano, E. Morschel, and S. Miyachi. Molecular structure, localization and function of biliproteins in the chlorophyll *a/d* containing oxygenic photosynthetic prokaryote *Acaryochloris marina*. *Biochim Biophys Acta*, 1412(3):250–261, 1999.

J. P. Huelsenbeck and F. Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755, 2001.

iWorx (CB Sciences). *Tech Note: ISE-730 Oxygen Electrode and DO2-100*

Current-to-Voltage adapter. 1 Washington Street, Suite 404, Dover, NH 03820, 2006.

Y. Kashiyama, H. Miyashita, S. Ohkubo, N. O. Ogawa, Y. Chikaraishi, Y. Takano, H. Suga, T. Toyofuku, H. Nomaki, H. Kitazato, T. Nagata, and N. Ohkouchi. Evidence of global chlorophyll *d*. *Science*, 321:658, 2008.

C. L. Kingsford, K. Ayanbule, and S. L. Salzberg. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol*, 8(2):R22, 2007.

J. L. Kirschvink, E. J. Gaidos, L. E. Bertani, N. J. Beukes, J. Gutzmer, L. N. Maepa, and R. E. Steinberger. Paleoproterozoic snowball earth: extreme climatic and geochemical global change and its biological consequences. *Proc Natl Acad Sci U S A*, 97(4):1400–1405, 2000.

G. Konopa and K. Taylor. Coliphage lambda ghosts obtained by osmotic shock or LiCl treatment are devoid of J- and H-gene products. *J Gen Virol*, 43(3):729–733, 1979.

E. V. Koonin. Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. *Biol Direct*, 1:39, 2006.

E. V. Koonin, A. R. Mushegian, and P. Bork. Non-orthologous gene displacement. *Trends Genet*, 12(9):334–336, Sep 1996.

M. Kühl, M. Chen, P. J. Ralph, U. Schreiber, and A. W. Larkum. Ecology: a niche for cyanobacteria containing chlorophyll *d*. *Nature*, 433(7028):820, 2005.

L. R. Kump. The rise of atmospheric oxygen. *Nature*, 451(7176):277–278, 2008.

J. Kwak and A. J. Bard. Scanning electrochemical microscopy. Theory of the feedback mode. *Anal Chem*, 61:1221–1227, 1989.

J. Labonté, K. E. Reid, and C. A. Suttle. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl Environ Microbiol*, 75: 3634–3640, 2009.

A. W. Larkum and M. Kühl. Chlorophyll *d*: the puzzle resolved. *Trends Plant Sci*, 10(8):355–357, 2005.

R. Lavigne, W. D. Sun, and G. Volckaert. PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics*, 20(5):629–635, 2004.

C. N. Lee, R. M. Hu, T. Y. Chow, J. W. Lin, H. Y. Chen, Y. H. Tseng, and S. F. Weng. Comparison of genomes of three *Xanthomonas oryzae* bacteriophages. *BMC Genomics*, 8:pp. –, 2007.

M. E. Levin, R. W. Hendrix, and S. R. Casjens. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol*, 234(1):124–139, 1993.

W. K. W Li. Composition of ultraphytoplankton in the central North-Atlantic. *Mar Ecol Prog Ser*, 122:1–8, 1995.

Z. Li and M. P. Deutscher. The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. *Proc Natl Acad Sci U S A*, 92(15): 6883–6886, 1995.

Z. Li, S. Pandit, and M. P. Deutscher. Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. *Rna*, 5(1):139–146, 1999.

D. Lindell, M. B. Sullivan, Z. I. Johnson, A. C. Tolonen, F. Rohwer, and S. W.

Chisholm. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A*, 101(30):11013–11018, 2004.

D. Lindell, J. D. Jaffe, Z. I. Johnson, G. M. Church, and S. W. Chisholm. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, 438(7064):86–89, 2005.

D. Lindell, J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen, W. R. Hess, G. M. Church, and S. W. Chisholm. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, 449(7158):83–86, 2007.

H. B. Liu, H. A. Nolla, and L. Campbell. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat Microb Ecol*, 12:39–47, 1997.

X. Liu, M. Shi, S. Kong, Y. Gao, and C. An. Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. *Virology*, 366(1):28–39, 2007.

X. Liu, S. Kong, M. Shi, L. Fu, Y. Gao, and C. An. Genomic analysis of freshwater cyanophage Pf-WMP3 infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb Ecol*, 56(4):671–680, 2008.

J. Loessner and S. Hagens. Chapter 13: Bacteriophages of Listeria. In H. Goldfine and H. Shen, editors, *Listeria Monocytogenes: Pathogenesis and Host Response*. Springer, 2007.

T. M. Lowe and S. R. Eddy. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 25(5): 955–964, 1997.

J. Lu, F. Chen, and R. E. Hodson. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl Environ Microbiol*, 67(7):3285–3290, 2001.

S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6):491–511, 1943.

E. Magnani, K. Sjolander, and S. Hake. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell*, 16(9):2265–2277, 2004.

T. Maniatis, E. F. Fritsch, and J. Sambrook. *Molecular Cloning: A Laboratory Manual, Volume 1*. Cold Spring Harbour Laboratory, USA, 1982a.

T. Maniatis, E. F. Fritsch, and J. Sambrook. *Molecular Cloning: A Laboratory Manual, Volume 3*. Cold Spring Harbour Laboratory, USA, 1982b.

J. Maniloff and H. W. Ackermann. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. *Arch Virol*, 143(10):2051–2063, 1998.

N. H. Mann. The third age of phage. *PLoS Biol*, 3(5):e182, 2005.

N. H. Mann, M. R. Clokie, A. Millard, A. Cook, W. H. Wilson, P. J. Wheatley, A. Letarov, and H. M. Krisch. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol*, 187(9):3188–3200, 2005.

W.M. Manning and H.H. Strain. Chlorophyll *d*, a green pigment of red algae. *J Biol Chem*, 151:1–19, 1943.

J. Marquardt, H. Senger, H. Miyashita, S. Miyachi, and E. Morschel. Isolation and characterization of biliprotein aggregates from *Acaryochloris marina*, a *Prochloron*-like prokaryote containing mainly chlorophyll *d*. *FEBS Lett*, 410(2-3):428–432, 1997.

Sophie Martin. *Photoelectrochemistry of immobilised photosynthetic components: from chlorophyll to intact chloroplasts*. PhD thesis, University of Warwick, 2007.

P. Mehta, K. Katta, and S. Krishnaswamy. HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. *Protein Sci*, 13:295–300, 2004.

A. Millard, K. Zwirglmaier, M. Downey, N. H. Mann, and D. J. Scanlan. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localised to a hyperplastic region: implications for mechanisms of cyanophage evolution. submitted. *Environ Microbiol*, Epub ahead of print:pp. unknown, 2009.

A. D. Millard. Isolation of cyanophages from aquatic environments. *Methods Mol Biol*, 501:33–42, 2009.

A. D. Millard and N. H. Mann. A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. *J Mar Biol Ass UK*, 86(3):507–515, 2006.

E. S. Miller, E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Rüger. Bacteriophage T4 genome. *Microbiol Mol Biol Rev*, 67:85–156, 2003.

S. R. Miller, S. Augustine, T. L. Olson, R. E. Blankenship, J. Selker, and A. M. Wood. Discovery of a free-living chlorophyll *d*-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. *Proc Natl Acad Sci U S A*, 102(3):850–855, 2005.

M. Mimuro, S. Akimoto, I. I. Yamazaki, H. Miyashita, and S. Miyachi. Fluorescence properties of chlorophyll *d*-dominating prokaryotic alga, *Acaryochloris marina*: studies using time-resolved fluorescence spectroscopy on intact cells. *Biochim Biophys Acta*, 1412(1):37–46, 1999.

M. Mimuro, K. Hirayama, K. Uezono, H. Miyashita, and S. Miyachi. Uphill energy transfer in a chlorophyll *d*-dominating oxygenic photosynthetic prokaryote, *Acaryochloris marina*. *Biochim Biophys Acta*, 1456(1):27–34, 2000.

M. Mimuro, S. Akimoto, T. Gotoh, M. Yokono, M. Akiyama, T. Tsuchiya, H. Miyashita, M. Kobayashi, and I. Yamazaki. Identification of the primary electron donor in PS II of the Chl *d*-dominated cyanobacterium *Acaryochloris marina*. *FEBS Lett*, 556(1-3):95–98, 2004.

S. Mirold, W. Rabsch, H. Tschape, and W. D. Hardt. Transfer of the *Salmonella* type III effector *sopE* between unrelated phage families. *J Mol Biol*, 312(1):7–16, 2001.

H. Miyashita, H. Ikemoto, N. Kurano, K. Adachi, M. Chihara, and S. Miyachi. Chlorophyll *d* as a major pigment. *Nature*, 383:402, 1996.

H. Miyashita, K. Adachi, K. Norihide, H. Ikemoto, M. Chihara, and S. Miyachi. Pigment composition of a novel oxygenic photosynthetic prokaryote containing chlorophyll *d* as the major chlorophyll. *Plant Cell Physiol*, 38 (3):274–281, 1997.

H. Miyashita, H. Ikemoto, N. Kurano, S. Miyachi, and M. Chihara. *Acaryochloris marina* gen. et sp. nov. (cyanobacteria), an oxygenic photosynthetic prokaryote containing Chl *d* as a major pigment. *J Phycol*, 39(6): 1247–1253, 2003.

R. Mole, D. Meredith, and D. G. Adams. Growth and phage resistance of *Anabaena* sp. strain PCC7120 in the presence of cyanophage AN-15. *J Appl Phycol*, 9:339–345, 1997.

L. R. Moore, G. Rocap, and S. W. Chisholm. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*, 393(6684):464–467, 1998.

L. R. Moore, A. Coe, E. R. Zinser, M. A. Saito, M. B. Sullivan, D. Lindell, K. Frois-Moniz, J. B. Waterbury, and S. W. Chisholm. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr: Methods 5*, 5: 353–362, 2007.

A. Murakami, H. Miyashita, M. Iseki, K. Adachi, and M. Mimuro. Chlorophyll *d* in an epiphytic cyanobacterium of red algae. *Science*, 303(5664): 1633, 2004.

D. Nelson. Phage taxonomy: we agree to disagree. *J Bacteriol*, 186(21): 7029–7031, 2004.

S. Ohkubo, H. Miyashita, A. Murakami, H. Takeyama, T. Tsuchiya, and M. Mimuro. Molecular detection of epiphytic *Acaryochloris* spp. on marine macroalgae. *Appl Environ Microbiol*, 72(12):7912–7915, 2006.

F. Oltmanns. Über die Kultur und Lebensbedingungen der Meersalgen. *Jahrb Wiss Botanik*, 23:349–440, 1892.

O. Ortega-Morales, J. Guezennec, G. Hernández-Duque, C. C. Gaylarde, and P. M. Gaylarde. Phototrophic biofilms on ancient Mayan buildings in Yucatan, Mexico. *Curr Microbiol*, 40(2):81–85, 2000.

L. Overbergh, D. Valckx, M. Waer, and C. Mathieu. Quantification of murine cytokine mRNAs using real time quantitative reverse transcriptase PCR. *Cytokine*, 11:305–312, 1999.

E. Pagaling, R. D. Haigh, W. D. Grant, D. A. Cowan, B. E. Jones, Y. Ma, A. Ventosa, and S. Heaphy. Sequence analysis of an archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. *BMC Genomics*, 8:410, 2007.

R. D. Page. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*, 12(4):357–358, 1996.

B. Palenik. Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol*, 67(2):991–994, 2001.

M. R. Parsek and Fuqua C. Biofilms 2003: emerging themes and challenges in studies of surface-associated microbial life. *J Bact*, 186(14):4427–4440, 2004.

F. Partensky, W. R. Hess, and D. Vaulot. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev*, 63 (1):106–127, 1999.

F. B. Perler. InBase, the intein database. *Nucleic Acids Res*, 30:383–384, 2002.

M. Pertea, S. M. Mount, and S. L. Salzberg. A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinformatics*, 8:159, 2007.

Z. Petrášek, F. J. Schmitt, C. Theiss, J. Huyer, M. Chen, A. Larkum, H. J. Eichler, K. Kemnitz, and H. J. Eckert. Excitation energy transfer from phycobiliprotein to chlorophyll *d* in intact cells of *Acaryochloris marina* studied by time- and wavelength-resolved fluorescence spectroscopy. *Photochem Photobiol Sci*, 4(12):1016–1022, 2005.

M. W. Pfaffl. Chapter 3: Relative quantification. In M. T. Dorak, editor, *Real-time PCR*, pages 63–82. Taylor & Francis, 2006.

W. H. Pope, P. R. Weigele, J. Chang, M. L. Pedulla, M. E. Ford, J. M. Houtz, W. Jiang, W. Chiu, G. F. Hatfull, R. W. Hendrix, and J. King. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a “horned” bacteriophage of marine *Synechococcus*. *J Mol Biol*, 368(4):966–981, 2007.

C. Proux, D. van Sinderen, J. Suarez, P. Garcia, V. Ladero, G. F. Fitzgerald, F. Desiere, and H. Brüssow. The dilemma of phage taxonomy illustrated

by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. *J Bacteriol*, 184(21):6026–6036, 2002.

Rank Brothers Ltd. *Digital Model 10 and Model 20 Controller. Operating Manual*. 56, High Street, Bottisham, Cambridge, CB5 9DA, England, issue 1a edition, February 2002.

T. Renger and E. Schlodder. The primary electron donor of photosystem II of the cyanobacterium *Acaryochloris marina* is a chlorophyll *d* and the water oxidation is driven by a chlorophyll *a*/chlorophyll *d* heterodimer. *J Phys Chem B*, 112(25):7351–7354, 2008.

R. Rippka. Isolation and purification of cyanobacteria. *Methods Enzymol*, 167:3–27, 1988.

R. Rippka, J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol*, 111:1–61, 1979.

R. Rippka, T. Coursin, W. Hess, C. Lichtle, D. J. Scanlan, K. A. Palinska, I. Iteman, F. Partensky, J. Houmard, and M. Herdman. *Prochlorococcus marinus* Chisholm et al. 1992 subsp. *pastoris* subsp. nov. strain PCC 9511, the first axenic chlorophyll *a*2/*b*2-containing cyanobacterium (Oxyphotobacteria). *Int J Syst Evol Microbiol*, 50 Pt 5:1833–1847, 2000.

R. J. Ritchie. Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res*, 89(1):27–41, 2006.

F. Ronquist and J. P. Huelsenbeck. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574, 2003.

S. Rozen and H. J. Skaletsky. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132:365–386, 2000.

D. B. Rusch, A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol*, 5(3):e77, 2007.

K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945, 2000.

S. L. Salzberg, A. L. Delcher, S. Kasif, and O. White. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res*, 26(2): 544–548, 1998.

A. Savitzky and M.J.E. Golay. Smoothing and differentiation of data by simplified least square procedures. *Anal Chem*, 36:1627–1639, 1964.

H. Scheer. *Chlorophylls*. CRC Press, Inc, 1991.

H. Schiller, H. Senger, H. Miyashita, S. Miyachi, and H. Dau. Light-harvesting in *Acaryochloris marina* - spectroscopic characterization of a chlorophyll *d*-dominated photosynthetic antenna system. *FEBS Lett*, 410(2-3):433–436, 1997.

V. Seguritan, I. W. Feng, F. Rohwer, M. Swift, and A. M. Segall. Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. *J Bacteriol*, 185(21):6434–6347, 2003.

R. Seshadri, S. A. Kravitz, L. Smarr, P. Gilna, and M. Frazier. CAMERA: a community resource for metagenomics. *PLoS Biol*, 5(3):e75, 2007.

J. Shan, Y. Jia, M. R. Clokie, and N. H. Mann. Infection by the 'photosynthetic' phage S-PM2 induces increased synthesis of phycoerythrin in *Synechococcus* sp. WH7803. *FEMS Microbiol Lett*, 283(2):154–161, 2008.

A. Shimada, M. Nishijima, and T. Maruyama. Seasonal abundance of *Prochlorococcus* in Suruga Bay, Japan in 1992-1993. *J Oceanogr*, 51: 289–300, 1995.

A. Shimada, T. Maruyama, and S. Miyachi. Vertical distributions and photosynthetic action spectra of two oceanic picophytoplankters, *Prochlorococcus marinus* and *Synechococcus* sp.˙ *Mar Biol*, 127(1):15–23, 1996.

L. L. Shipman, T. M. Cotton, J. R. Norris, and J. J. Katz. An analysis of the visible absorption spectrum of chlorophyll *a* monomer, dimer and oligomers in solution. *J Am Chem Soc*, 98:8222–8230, 1976.

D. A. Shub, H. Goodrich-Blair, and S. R. Eddy. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem Sci*, 19(10):402–404, 1994.

T. E. Shutt and M. W. Gray. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet*, 22(2):90–95, 2006a.

T. E. Shutt and M. W. Gray. Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *J Mol Evol*, 62(5):588–599, 2006b.

Sigma-Aldrich Co. *The Quantitative PCR Technical Guide*. 3050 Spruce St., St. Louis, MO 63103, 2008.

C. Six, J. C. Thomas, B. Brahamsha, Y. Lemoine, and F. Partensky. Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat Microb Ecol*, 35:17–29, 2004.

S. Spinelli, V. Campanacci, S. Blangy, S. Moineau, M. Tegoni, and C. Cambillau. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages: the RBP structure of the temperate phage TP901-1. *J Biol Chem*, 281(20):14256–14262, 2006.

G. S. Stent. The operon: on its third anniversary. *Science*, 144:816–820, 1964.

A. Sulakvelidze, Z. Alavidze, and J. G. Morris Jr. Bacteriophage therapy. *Antimicrob Agents Chemother*, 45(3):649–659, 2001.

M. B. Sullivan, J. B. Waterbury, and S. W. Chisholm. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature*, 424(6952):1047–1051, 2003.

C. A. Suttle. Viruses in the sea. *Nature*, 437(7057):356–361, 2005.

C. A. Suttle. Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol*, 5(10):801–812, 2007.

C. A. Suttle and A. M. Chan. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser*, 92:99–109, 1993.

B. E. Suzek, M. D. Ermolaeva, M. Schreiber, and S. L. Salzberg. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics*, 17(12):1123–1130, 2001.

W. D. Swingley, M. Chen, P. C. Cheung, A. L. Conrad, L. C. Dejesa, J. Hao, B. M. Honchak, L. E. Karbach, A. Kurdoglu, S. Lahiri, S. D. Mastrian,

H. Miyashita, L. Page, P. Ramakrishna, S. Satoh, W. M. Sattley, Y. Shimada, H. L. Taylor, T. Tomo, T. Tsuchiya, Z. T. Wang, J. Raymond, M. Mimuro, R. E. Blankenship, and J. W. Touchman. Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proc Natl Acad Sci U S A*, 105(6):2005–2010, 2008.

N. Tandeau de Marsac. Occurrence and nature of chromatic adaptation in Cyanobacteria. *J Bacteriol*, 130(1):82–91, 1977.

A. Telenti, P. Imboden, F. Marchesi, D. Lowrie, S. Cole, M. J. Colston, L. Matter, K. Schopfer, and T. Bodmer. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet*, 341(8846):647–650, 1993.

J. D. Thompson, T. J. Gibson, and D. G. Higgins. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*, Chapter 2:Unit 2 3, 2002.

G. Toledo and B. Palenik. *Synechococcus* diversity in the California current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. *Appl Environ Microbiol*, 63(11):4298–4303, 1997.

F. Torrella and R. Y. Morita. Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon: ecological and taxonomical implications. *Appl Environ Microbiol*, 37 (4):774–778, 1979.

M. Tsionsky, Z. G. Cardon, A. J. Bard, and R. B. Jackson. Photosynthetic electron transport in single guard cells as measured by scanning electrochemical microscopy. *Plant Physiol*, 113(3):895–901, 1997.

M. J. W. Veldhuis, G. W. Kraay, J. D. L. VanBleijswijk, and M. A. Baars. Seasonal and spatial variability in phytoplankton biomass, productivity and

growth in the north-western Indian Ocean: the southwest and northeast monsoon, 1992-1993. *Deep Sea Res Part I-Oceanographic Res Pap*, 44: 425–449, 1997.

I. Vinga, A. Dröge, A. C. Stiege, R. Lurz, M. A. Santos, R. Daugelavičius, and P. Tavares. The minor capsid protein gp7 of bacteriophage SPP1 is required for efficient infection of *Bacillus subtilis*. *Mol Microbiol*, 61(6): 1609–1621, 2006.

M. K. Waldor and J. J. Mekalanos. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, 272:1910–1914, 1996.

K. Wang and F. Chen. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol*, 10(2):300–312, 2008.

T. S.-F. Wang. *DNA replication in eukaryotic cells. Chapter 15: Cellular DNA Polymerases*. Cold Spring Harbor Monograph Series. Cold Spring Harbor Laboratory Press, New York, 1996.

J. B. Waterbury, S. W. Watson, F. W. Valois, and D. G. Franks. Biological and ecological characterisation of the marine unicellular cyanobacterium *Synechococcus*. *Can Bull Fish Aquat Sci*, 214:71–120, 1986.

P. R. Weigele, W. H. Pope, M. L. Pedulla, J. M. Houtz, A. L. Smith, J. F. Conway, J. King, G. F. Hatfull, J. G. Lawrence, and R. W. Hendrix. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol*, 9(7):1675–1695, 2007.

N. J. West, W. A. Schonhuber, N. J. Fuller, R. I. Amann, R. Rippka, A. F. Post, and D. J. Scanlan. Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology*, 147(Pt 7):1731–1744, 2001.

J. P. Wilburn, D. W. Wright, and D. E. Cliffel. Imaging of voltage gated alamethicin pores in a reconstituted bilayer lipid membrane via scanning electrochemical microscopy. *Analyst*, 131(2):311–316, 2006.

W. H. Wilson. *Characterisation of viruses infecting marine phytoplankton*. PhD thesis, University of Warwick, 1994.

W. H. Wilson, I. R. Joint, N. G. Carr, and N. H. Mann. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol*, 59(11):3736–3743, 1993.

W. H. Wilson, N. G. Carr, and N. H. Mann. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. *J Phycol*, 32:506–516, 1996.

W. B. Wood. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol*, 16(1):118–133, 1966.

M. Wyman, R.P.F. Gregory, and N.G. Carr. Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science*, 230(4727): 818–820, 1985.

J. Xu, R. W. Hendrix, and R. L. Duda. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell*, 16:11–21, 2004.

T. Yasukawa, Y. Kondo, I. Uchida, and T. Matsue. Imaging of cellular activity of single cultured cells by scanning electrochemical microscopy. *Chem Lett*, 27(8):767, 1998.

J. Yin. A quantifiable phenotype of viral propagation. *Biochem Biophys Res Commun*, 174(2):1009–1014, 1991.

J. Yin. Evolution of bacteriophage T7 in a growing plaque. *J Bacteriol*, 175 (5):1272–1277, 1993.

S. Yooseph, G. Sutton, D. B. Rusch, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C. S. Miller, H. Li, S. T. Mashiyama, M. P. Joachimiak, C. van Belle, J. M. Chandonia, D. A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B. J. Raphael, V. Bafna, R. Friedman, S. E. Brenner, A. Godzik, D. Eisenberg, J.E. Dixon, S. S. Taylor, R. L. Strausberg, M. Frazier, and J. C. Venter. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol*, 5(3):e16, 2007.

T. Yoshida, K. Nagasaki, Y. Takashima, Y. Shirai, Y. Tomaru, Y. Takao, S. Sakamoto, S. Hiroishi, and H. Ogata. Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol*, 190(5):1762–1772, 2008.

Y. Zhong, F. Chen, S. W. Wilhelm, L. Poorvin, and R. E. Hodson. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Appl Environ Microbiol*, 68(4):1576–1584, 2002.

Y. Zuo and M. P. Deutscher. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res*, 29(5):1017–1026, 2001.

M. Zweig and D. J. Cummings. Structural proteins of bacteriophage T5. *Virology*, 51(2):443–453, 1973.